



## **FACULTY OF GRADUATE STUDIES**

### **Using Molecular Modeling to Identify Zinc Finger Protein-DNA Recognition Patterns: A Study of Four Finger Proteins and Dimer Binding to DNA**

استخدام التمثيل الجزيئي لتحديد أنماط تعرف بروتين أصبع الزنك على الحمض النووي: دراسة ارتباط أربعة أنماط من بروتين أصبع الزنك والبروتين الثنائي مع الحمض النووي

This thesis is submitted in partial fulfillment of the requirements for the master's Degree in Applied Chemistry from the Faculty of Graduate Studies at Birzeit University, Palestine

By: Reema Mazen Siam

Supervisor: Dr. Mazen Hamed

Sep/2017

Using Molecular Modeling to Identify Zinc Finger  
Protein-DNA Recognition Patterns: A Study of Four  
Zinc Finger Proteins and their Dimer Binding to DNA

By

Reema Mazen Siam

This thesis was defended successfully on the 7<sup>th</sup> of September, 2017 and approved  
by:

Committee Members

Signature

Dr. Mazen Hamed, Ph.D

.....

(Principal Advisor)

Wael Karain, Ph. D

.....

(Member of thesis committee)

Abdallah Sayyed-Ahmad, Ph. D

.....

(Member of thesis committee)

# ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Mazen Hamed for his continuous guidance, generous support, and motivation through the project and for sharing his knowledge.

I would also like to thank my mother for her continues motivation and support without which I could not achieve my goal.

Finally, I would like to express my gratitude to my family and my husband for providing me with continuous support and encouragement throughout my years of study and through the process of researching and writing this thesis.

# Table of Contents

ACKNOWLEDGEMENTS.....	I
ABSTRACT.....	VI
ملخص.....	VII
List of Figures.....	VIII
List of Tables.....	XIII
INTRODUCTION.....	1
1.1 Zinc Finger Proteins.....	1
1.1.1 C <sub>2</sub> H <sub>2</sub> Zinc Fingers.....	2
1.1.1.1 Folding of C <sub>2</sub> H <sub>2</sub> Zinc Fingers.....	4
1.1.1.2 Tandem of C <sub>2</sub> H <sub>2</sub> Zinc Fingers.....	5
1.1.1.3 Features of the Binding Free Energy of C <sub>2</sub> H <sub>2</sub> Zinc Fingers to their Target DNA.....	5
1.1.1.4 Three Zinc Finger Proteins versus Multiple Zinc Finger Proteins binding to DNA...	6
1.1.1.5 Human PRDM9 allele-A ZnF Domain bound to DNA Sequence II.....	8
1.1.2 Zinc Cluster Protein Zn <sub>2</sub> Cys <sub>6</sub> (Binuclear Cluster).....	10
1.1.2.1 Structure - Function Relationship of Zinc Cluster Proteins.....	10
1.1.2.2 Specificity of the DNA-Binding Process.....	13
1.1.2.3 PPR1-DNA Complex.....	14
1.1.2.4 PUT3-DNA Complex.....	15
1.1.3 Biological Applications of Zinc Finger Proteins.....	17
1.2.1 Binding Free Energy Calculations.....	19
1.2.1 MM_GBSA Method for Predicting the binding of Protein to DNA.....	20
COMPUTATIONAL METHODS.....	25
2.1 Initial Structures of Protein-DNA Complexes.....	25
2.2 Equilibration of the solvated systems.....	26
2.3 Analysis of Output Files from Equilibration Process.....	27
2.4 Production Simulations of the Solvated Systems.....	30
2.5 Calculating the Binding Free Energy of Protein-DNA complexes.....	30
2.6 Calculation of the Entropy Contribution.....	31
2.7 Analysis of Results.....	31
RESULTS AND DISCUSSION.....	32

3.1 Protein-DNA Interaction .....	32
3.1.1 Binding of PRDM9 <sub>A</sub> and Its Alleles to DNA .....	32
3.1.1.1 Overall Structure of PRDM9 <sub>A</sub> -DNA Complex .....	33
3.1.1.2 Binding Affinities of PRDM9 Alleles to DNA .....	35
3.1.1.3 Binding of PRDM9 <sub>A</sub> Compared to the three Finger protein 1AAY to DNA .....	41
3.1.1.4 Alleles of PRDM9 Protein Binding to DNA .....	42
3.1.1.5 PRDM9 <sub>A</sub> - DNA Interaction: Hydrogen-Bonding .....	46
3.1.2 Binding of PPR1 and Its Mutants to DNA .....	55
3.1.2.1 Overall Structure of PPR1-DNA Complex .....	55
3.1.2.2 Binding of PPR1 Compared to the three Zinc Finger Protein 1AAY to DNA .....	58
3.1.2.3 Binding Affinities of PPR1-DNA Complex and its Mutant.....	61
3.1.2.4 PPR1- DNA Interaction: Hydrogen-Bonding.....	62
3.1.3 Binding of PUT3 and Its Mutants to DNA.....	66
3.1.3.1 Overall Structure of PUT3-DNA Complex .....	67
3.1.3.2 Binding of PUT3 Compared to the three Zinc Finger Protein 1AAY to DNA .....	68
3.1.3.3 Comparison of PUT3-DNA Complex with PPR1-DNA Complex (Binuclear Cluster) .....	70
3.1.3.4 Binding Affinities of PUT3 protein to DNA and its Mutants to DNA.....	71
3.1.3.5 PUT3- DNA Interaction: Hydrogen-Bonding .....	74
3.2 The Binding Free Energy of Zinc Finger-DNA Complexes: MM/GBSA Versus MM/PBSA.....	77
CONCLUSION.....	80
APPENDICES .....	83
REFERENCES .....	125

## ABSTRACT

Binding of Zinc finger proteins to their DNA targets is an important process for regulating many biological processes. In this study, four finger and dimers ZnF proteins bound to their target DNA are simulated using Molecular dynamics simulations to obtain their equilibrium structures and extract important interactions determining these zinc finger proteins-DNA binding specificity.

Furthermore, binding free energies of these zinc finger proteins to their DNA targets are calculated using MM/GBSA and MM/PBSA approaches. In particular, the multidomain zinc finger (ZnF) protein PRDM9 (PDB ID: 5EGB) and its variants L9/L24, L13, L20 bind AT, and L20 bind GC calculated binding free energies were -29.48, -31.92, -27.83, -22.44, -9.97 kcal/mol, respectively. While calculated binding free energies of dimer transcription factors PPR1 (PDB ID: 1PYI) and PUT3 (PDB ID: 1ZME) to their target DNA were -7.35 and -15.79 kcal/mol. The aforementioned calculated binding free energies showed low correlation with experimentally obtained energies ( $R = 0.69$  and  $0.59$  for MM/PBSA and MM/GBSA, respectively).

## ملخص

عملية ارتباط بروتينات أصبع الزنك مع الحمض النووي تعد من العمليات الهامة لتنظيم العديد من العمليات البيولوجية. في هذه الدراسة، تم استخدام حسابات التمثيل الجزيئي لحساب طاقة الربط لأربعة أنماط من أصبع الزنك والبروتين الثنائي للحصول على المركبات في حالة الاتزان ولاستخراج الروابط المهمة التي تحدد الخصوصية لبروتينات أصبع الزنك المرتبطة مع الحمض النووي.

علاوة على ذلك، طاقة الربط لبروتينات أصبع الزنك مع الحمض النووي تم حسابها باستخدام طريقة (Molecular Mechanics Poission Boltzman (Generalized born) Surface Area). على

وجه الخصوص، بروتين أصبع الزنك المتعدد (PRDM9<sub>A</sub>, PDB ID:5EGB) ومتغيراته (L9/24, L13, L20 bind CA, and L20 bind CG) قيم طاقة الربط الناتجة هي: -29.48, -31.92, -27.83, -22.44, and -9.97 kcal/mol) بينما قيم طاقة الربط للبروتينات الثنائية PPR1 (PDB ID: 1PYI) and PUT3 (PDB ID: 1ZME) مع الحمض النووي هي -7.35, -15.79 kcal/mol). أظهرت طاقات الربط المذكورة ارتباطا ضعيفا مع الطاقات التي تم الحصول عليها تجريبيا (MM/PBSA ل R = 0.69 و MM/GBSA ل R = 0.59).

# List of Figures

Figure 1.1: Primary structure of C <sub>2</sub> H <sub>2</sub> zinc finger motifs in (a) prokaryotes, (b) green plants, and animals (c).....	3
Figure 1.2: Structure of ZnF in TFIIA. Zinc ion (yellow), cysteine (green) and histidine (blue)...	3
Figure 1.3: Zinc interaction motivates the proper DNA binding conformation of the C <sub>2</sub> H <sub>2</sub> finger polypeptide.....	4
Figure 1.4: The Zif268-DNA complex. A) Major groove binding. Fingers 1, 2, and 3 of Zif268 are red, yellow and purple, DNA is blue. On the right is color coded parallel to finger color on the left. B) The side chain–base interactions in the Zif268 complex.....	7
Figure 1.5: Domain organization of the hPRDM9 protein.....	9
Figure 1.6: Crystal structure of hPRDM9A ZnF8–12 in complex with the THE1B recombination hot spot sequence. ZnF8–11(color-coded blue [ZnF8], cyan [ZnF9], orange [ZnF10], and magenta [ZnF11]) are shown in cartoon.....	9
Figure 1.7: The Zn <sub>2</sub> Cys <sub>6</sub> zinc cluster motif. A) The amino acid sequence of the Gal4p Zn <sub>2</sub> Cys <sub>6</sub> zinc cluster motif. B) The binding of zinc ions facilitates the tight folding of the Zn <sub>2</sub> Cys <sub>6</sub> motif.....	10
Figure 1.8: Functional domains of zinc cluster.....	11
Figure 1.9: Zinc ion coordination in zinc cluster proteins.....	11
Figure 1.10: Crystal structures of the DBDs of some Zn(II)2Cys6 regulators. Yellow spheres correspond to zinc atoms.....	12
Figure 1.11: A model for zinc cluster protein DNA recognition.....	13
Figure 1.12: PPRI DNA-binding domain.....	14
Figure 1.13: Structure of Zn domain bound to DNA. a) The interaction involves Zn domain. Protein is blue, the amino-acid side chains is green that interact with DNA phosphate oxygens are red. Protein residues that make contacts to functional groups on the DNA bases are yellow. Broken lines represent hydrogen bonds between protein and	



DNA bases. b) Schematic diagram showing all phosphate and base contacts with protein. Van der Waals contacts are representing with broken.....	15
Figure 1.14: PUT3 DNA-binding domain and its DNA binding sites. a, Amino acid sequence of the PUT3 in the crystallographic analysis. b, 16-bp PUT3 binding site from the PUT2.....	16
Figure 1.15: Overall structure of PUT3- DNA complex. DNA (red), CGG triplet (purple) metal ions (yellow).....	17
Figure 1.16: Different regulations of ZNF proteins' functions in cancer progression.....	19
Figure 1.17: Binding free energy prediction for a solvated system as the difference between bound free energy and unbound free energy end states. Solvent shell is light blue..	20
Figure 1.18: Thermodynamic cycle demonstrating steps of calculating binding free energy using MM/GBSA method. Solvent shell is light blue, vacuum is black.....	21
Figure 2.1: Zinc ion coordination in zinc cluster proteins.....	25
Figure 2.2: Plots against time for the heating and equilibration phases of the density (a), temperature (b), pressure (c), and energy (d).....	28
Figure 3.1 : Backbone (CA, C, N) RMSD vs. Frame for the production phase of the MD refinement for PRDM <sub>A</sub> -DNA complex.....	32
Figure 3.2 : Overall structure of PRDM <sub>9A</sub> -DNA complex after molecular modeling simulation, showing the four zinc fingers. A) ZnF1-4 in complex with the DNA strands , ZnF1-4 are represented as a light brown and the coordinated zinc ions are represented as blue spheres. B)The PRDM <sub>9</sub> complex are represented as a rounded ribbon, ZnF1-4 are red and the coordinated zinc ions are represented as a blue spheres. C) The PRDM <sub>9</sub> complex are represented as a hydrophobisity surface.....	33
Figure 3.3: The change in the absolute energy ( $\Delta G_{\text{absolute}}$ ) for different alleles of PARMD9. Binding energy ( $\Delta H$ ) and entropy contribution ( $T\Delta S$ ) to the absolute energy from both MM/GBSA and MM/PBSA. Absolute energy change for $\Delta G_{\text{MM/GBSA}}$ labeled green and for $\Delta G_{\text{MM/PBSA}}$ orange for different alleles. Binding energy $\Delta H_{\text{MM/GBSA}}$	

labeled blue and $\Delta H_{MM/PBSA}$ purple) . Entropy contribution ( $T\Delta S$ ) labeled red...	39
Figure 3.4: Change in the binding energy for different alleles of PARMD9. Binding energy ( $\Delta H_{MM/GBSA}$ and $\Delta H_{MM/PBSA}$ ) are represented as blue and purple, respectively. $G_{ELE}$ is represented as a red squares with red line, VdW energy is represented as a green triangles with green line. Electrostatic energy multiplied by $10^2$ .....	40
Figure 3.5: Correlation between binding energy ( $\Delta H_{MM/GBSA}$ and $\Delta H_{MM/PBSA}$ ) and electrostatic energy ( $G_{ELE}$ ) and Van der Waals (VdW)of amino acid residues upon different alleles of PARMD9. A) the $r$ values of 0.91 indicates a strong correlation between the binding energy and electrostatic energy. B) The $r$ values of both $G_{ELE}$ and VdW are almost the same so that indicate a strong correlation between binding energy and both electrostatic energy and Van der Waals.....	40
Figure 3.6: The non-A alleles of hPRDM9 where a single amino acid substitution at Znf3 and Znf4 resulted in alleles L9/L24, L20 or L13, respectively.....	44
Figure 3.7 : A) DNA strand (5`-3`) of the PARMD9 <sub>L20</sub> -DNA complex. B) Mutant DNA strand (5`-3`) of the PARMD9 <sub>L20</sub> -DNA complex. ....	44
Figure 3.8: 5EGB-DNA contacts: Guanine H-bond contacts with: A) Arg63, B) Arg91 and C) Arg147 residues at position 6 with inter-atomic distances in F1, F2, F4.....	47
Figure 3.9: 5EGB-DNA contacts: Guanines H-bond contacts with: A) His60, B) His88 and C) His144 residues at position 3 with interatomic distances in ZNF1, 2, 4.....	48
Figure 3.10: 5EGB-DNA contacts: Adenine and Thymine H-bond contacts with Asn116 residues at position 3 in F3: A) Adenine contact and B) Thymine water mediated nonspecific contact.....	49
Figure 3.11: Backbone (CA, C, N) RMSD vs. Frame for the production phase of the MD refinement for 1PYI.....	55
Figure 3.12: Overall structure of PPR1-DNA complex after molecular modeling simulation, showing the two binuclear cluster ( $Zn_2Cys_6$ ). A) PPR1 complex with the DNA strands , two binuclear cluster are represented as a light brown and the coordinated	

zinc ions are represented as blue spheres. B)The PPR1-DNAcomplex are represented as a rounded ribbon, first binuclear cluster are red , the second one are green the coordinated zinc ions are represented as a blue spheres. C) The PPR1 complex are represented as a hydrophobisity surface, DNA strands are a light brown.....56

Figure 3.13: A) Amino acid sequence of the Zn domain of the PPR1. B) Structure of the Zn domain by Chimera showing two short helices and the coordinated zinc ions as blue spheres.....57

Figure 3.14: The Zif268-DNAcomplex showing the three zinc fingers of Zif268 bound in the major groove of the DNA. Fingers are spaced 3-base pairs intervals. The DNA is blue; fingers 1, 2, and 3 of Zif268 are red, yellow, and purple, respectively; and the coordinated zinc ions are represented as silver spheres. The DNA sequence of Zif268 site on the right is color-coded to indicated base contacts made by each finger.....59

Figure 3.15: The Binuclear cluster region of PPR1 after molecular dynamic simulations, showing the coordinated zinc ions as blue spheres and the Lys residue at position 40 and 41 in both nuclear clusters that make a series of interactions with the CGG bases near each end of the DNA as light brown.....59

Figure 3.16: PPR1-Binding sites containing two highly CGG elements near each end of the DNA. A) without mutation. B) with mutation at the CGG half sites.....61

Figure 3.17: Schematic diagram showing all phosphate and base contacts between 1PPR1 protein and DNA strand.....63

Figure 3.18: Backbone (CA, C, N) RMSD vs. Frame for the production phase of the MD refinement for 1ZME.....66

Figure 3.19: Overall structure of PUT3-DNA complex after molecular modeling simulation, showing the two binuclear cluster ( $Zn_2Cys_6$ ). A) PUT3 complex with the DNA strands , two binuclear cluster are represented as a light brown and the coordinated zinc ions are represented as blue spheres. B)The PUT3-DNAcomplex are represented as a rounded ribbon, first binuclear cluster are red , the second one are green the

coordinated zinc ions are represented as a blue spheres. C) The PUT3 complex are represented as a hydrophobisity surface, DNA strands are a light brown.....	67
Figure 3.20: The base contact of the individual fingers of 1AAY-DNA complex. Individual base are represented by rectangles, the primary strand of the DNA in the left and the secondary strand on the right. Shaded rectangles represents the 3 bases in the primary strand that are contacted by each finger. ....	69
Figure 3.21: The Binuclear cluster region of PUT3 after molecular dynamic simulations showing the residues at position 42 (Arg), and 44 (Arg) in both nuclear clusters are responsible for the contacts to the CGG end of the DNA .....	70
Figure 3.22: A) Complementary chain (3`-5`) of the PUT3-DNA complex. B) Mutant complementary chain (3`-5`) of the PUT3-DNA complex.....	72
Figure 3.23: Schematic representation of the 5IU unit.....	72
Figure 3.24: The correlation between the binding free energies calculated by A) MM/GBSA, B) MM/PBSA methods and the experimental values. ....	78
Figure A.1: Loading configurations file needed for AMBER ff14SB force field.....	83
Figure A.2: preparing <i>Xleap</i> to load the 5EGB, PPR1 and PUT3 PDB files.....	84
Figure A.3: Graphical representation of 5Iu unit.....	85
Figure A.4: The editor window of <i>Xleap</i> showing the graphical representation of bonded 5egb_dry.pdb.....	86
Figure A.5: Neutralization of 5egb complex through the addition of sodium ions.....	87
Figure A.6: A graphical representation of 5EGB solvated system.....	88
Figure A.7: The PyMOL viewer window showing how the amino Lysine (K) at position 114 in 5EGB-DNA complex is mutated by choosing Glutamic acid from the mutagenesis list after clicking “No mutation” option.....	90
Figure A.8: The UCSF Chimera window showing how the nucleotide adenine (A) at position 12 in chain D in 5EGB-DNA complex is mutated to guanine (G) by written “swapna G :12.d” in the command window.....	92

## List of Tables

Table 1.1: Thermodynamics parameter for zinc finger reaction.....	6
Table 3.1. Calculated energies (kcal/mol) for 1AAY, and PRDM <sub>9</sub> -DNA with different alleles complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM-GBSA and MM-PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting T $\Delta S$ from <i>nmode</i> calculations from the binding free energy.....	35
Table 3.2. Calculated energies for 1AAY, and the variant alleles of PRDM <sub>9</sub> -DNA complex using MM/GBSA. The Van der Waals energy (VdW), electrostatic energy( $G_{ELE}$ ), the polar energy ( $E_{GB}$ ), and the non-polar energy ( $E_{SURF}$ ) are shown.....	36
Table 3.3. Calculated energies for 1AAY, and the variant alleles of PRDM <sub>9</sub> -DNA complex using MM/PBSA. The Van der Waals energy (VdW), electrostatic energy( $G_{ELE}$ ), the polar energy ( $E_{PB}$ ), and the non-polar energy ( $E_{nopolar}$ ) are shown.....	37
Table 3.4: Correlation coefficients ( $r$ ) for different energy contributions to the total free energy of binding ( $\Delta H$ ) in both MM/GBSA and MM/PBSA methods .....	39
Table 3.5. Comparison of theoretically calculated $K_d$ values from both MM/GBSA and MM/PBSA with the experimentally derived values using Gel-shift assays. ....	45
Table 3.6. Calculated energies for PRDM <sub>9A</sub> -DNA complex and its different alleles. ....	45
Table 3.7. The direct contacts between 5EGB-DNA complex.....	50
Table 3.8. Calculated energies (kcal/mol) for 1AAY and 1ZME DNA complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM/GBSA and MM/PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting T $\Delta S$ from <i>nmode</i> calculations from the binding free energy.....	60
Table 3.9. Comparison of theoretically calculated $K_d$ values from both MMGBSA and MMPBSA with the experimentally derived values using Gel-shift assays.....	60
Table 3.10. The direct contacts between both binuclear clusters C <sub>34</sub> K <sub>35</sub> R <sub>36</sub> C <sub>37</sub> R <sub>38</sub> L <sub>39</sub> K <sub>40</sub> K <sub>41</sub> and the 5`-3` sequence and its 3`-5` complementary sequence. The numbering of protein	

sequence is less by one than the X-ray structural (K40 = K39 in our case).....63

Table 3.11. Calculated energies (kcal/mol) for 1AAY and 1ZME-DNA complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM/GBSA and MM/PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting  $T\Delta S$  from *nmode* calculations from the binding free energy.....73

Table 3.12. Comparison of theoretically calculated  $K_d$  values from both MM/GBSA and MM/PBSA with the experimentally derived values using Gel-shift assays.....73

Table 3.13. The direct contacts between binuclear clusters and the 5`-3` sequence and its 3`-5` complementary sequence. The numbering of protein sequence is more by four than the X-ray structural (R40 = R44 in our case).....74

Table 3.14: Calculated and experimental binding free energies (kcal/mol) using MM/GBSA and MM/PBSA methods for our protein-DNA complexes.  $G_{ELE}$  is multiplied by  $10^2$ .....79

# Chapter 1

## INTRODUCTION

DNA-binding proteins play a major role in the development of organisms by regulating many vital biological processes such as DNA replication and repair<sup>1,2</sup>.

There are several types of DNA-binding proteins such as helix-turn-helix, Zn-binding domains and leucine-zipper<sup>3</sup>. Many DNA-binding proteins recognize specific DNA sites either in the major groove or minor groove through small discrete domains which expose different designs to a prominent surface, a flexibly extended structure, or both so as to contact DNA specifically or nonspecifically<sup>1,3,4</sup>.

### 1.1 Zinc Finger Proteins

Zinc metal ion ( $\text{Zn}^{2+}$ ) is an abundant and nontoxic transition metal ion. It has unique properties when compared to other transition metal ions. The electron configuration of  $\text{Zn}^{2+}$  affects the chemistry in biological systems in four ways; 1) there is no ligand field stabilization energy when it coordinates the ligands and adopts coordination numbers four, five and six in protein environment. 2)  $\text{Zn}^{2+}$  can interact strongly with different types of ligands such as sulfur in cysteine, nitrogen from histidine and oxygen from aspartate, glutamate and water. 3) Contrary to copper and iron ions,  $\text{Zn}^{2+}$  is not redox active and does not form a precipitate in

oxidation-reduction reactions. 4)  $Zn^{2+}$  is labile thus the uptake and release of zinc metal become easier<sup>5</sup>.

Zinc finger proteins are DNA-binding motifs found in eukaryotic cells<sup>6,7</sup>. Their functions include DNA recognition, RNA packaging, transcriptional activation and protein folding<sup>6,8</sup>.

Xenopus (TFIIIA) the first zinc finger recognized as a repeated Zinc-binding motif which revealed a series of nine zinc finger domains. Each domain has a sequence of the form  $X_3\text{-Cys-}X_{2-4}\text{-Cys-}X_{12}\text{-His-}X_{3-4}\text{-His-}X_4$ , where X is an amino acid<sup>5,6,8</sup>. Since the discovery of TFIIA many zinc finger proteins were discovered such as the yeast gene ADRI, Wilms tumor gene (WT1) in human and Zif268 in mice<sup>5</sup>.

Zinc finger proteins composed of one or more zinc finger domains and each domain contains one or more zinc ion<sup>9</sup>. However, there are a different structural classes of zinc finger proteins according to the protein folding around the zinc ion. For example,  $Cys_2His_2$ ,  $Zn_2/Cys_6$  fingers<sup>8,10,11</sup>.

### **1.1.1 $C_2H_2$ Zinc Fingers**

$C_2H_2$  zinc finger proteins are considered one of the most common group of zinc finger proteins, which were First discovered in the *Xenopus laevis* transcription factor IIIA<sup>8,12,13</sup>. These proteins are found in prokaryotes and eukaryotes and humans (ZnF proteins)<sup>14</sup> (Figure1.1).



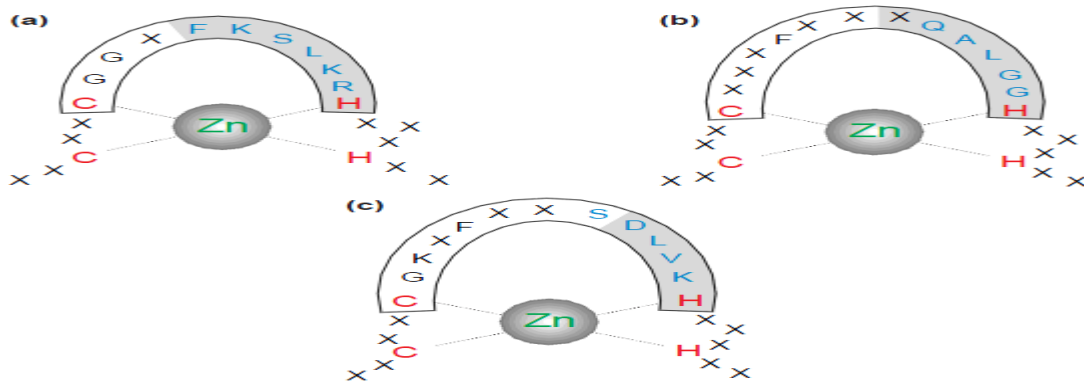


Figure 1.1: Primary structure of C<sub>2</sub>H<sub>2</sub> zinc finger motifs in (a) prokaryotes, (b) green plants, and animals (c)<sup>14</sup>.

A finger consists of a helix containing 20-30 amino acid residues. Its structure is stabilized by a tetrahedral zinc ion bound to two cysteine and two histidine residues<sup>8,12</sup>(Figure 1.2). The role of these fingers is predominant in protein-DNA, protein-RNA and protein-protein interactions<sup>15,16</sup>.

C<sub>2</sub>H<sub>2</sub> zinc finger are often described as X<sub>2</sub>CX<sub>2-4</sub>CX<sub>12</sub>HX<sub>2-8</sub> H pattern, where X is the amino acid residue, and the 12 residue region between the second cysteine and the first histidine most often has the pattern -X<sub>3</sub>-(F/Y)-X<sub>5</sub>-Ψ-X<sub>2</sub>- , where Ψ is a hydrophobic residue<sup>8,12,17</sup>.

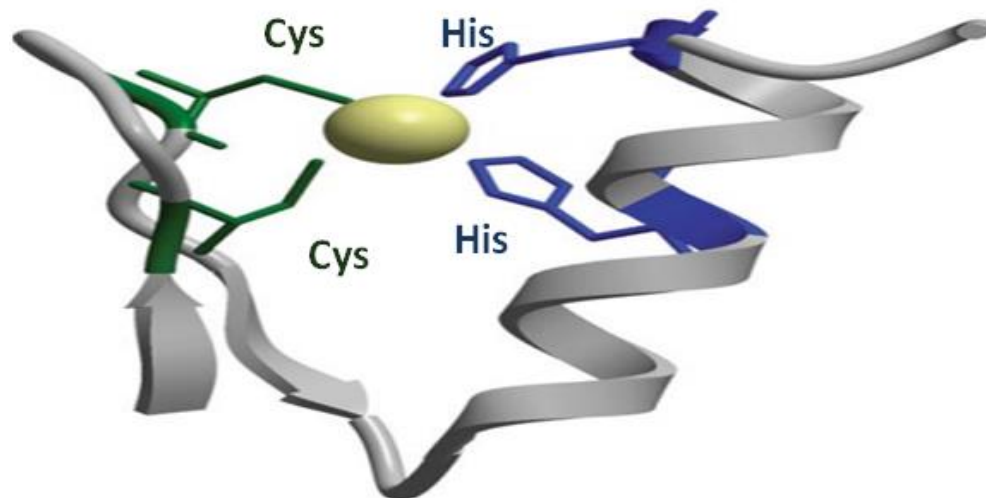


Figure 1.2: Structure of ZnF in TFIIA. Zinc ion (yellow), cysteine (green) and histidine (blue)<sup>17</sup>.

### 1.1.1.1 Folding of C<sub>2</sub>H<sub>2</sub> Zinc Fingers

Zinc finger protein has three states of folding. First is the unfolded state where the zinc finger cannot bind to DNA. Second is the folded state where the finger can bind to the DNA. The third is the DNA-bound form. The C<sub>2</sub>H<sub>2</sub> ZnF protein is active when Zn<sup>2+</sup> binds to two cysteine and two histidine residues<sup>8</sup>. The folded state has a lower free energy than that of the unfolded one mainly due to enthalpy change of -8.8 kcal/mol<sup>18</sup>(Figure 1.3).

Zinc ion interacts with the two cysteine residues separated by 2-4 amino acid residues and then with the two histidine residues. The zinc ion role is to help in forming specific contacts with the DNA bases<sup>8,12,19,20</sup>.

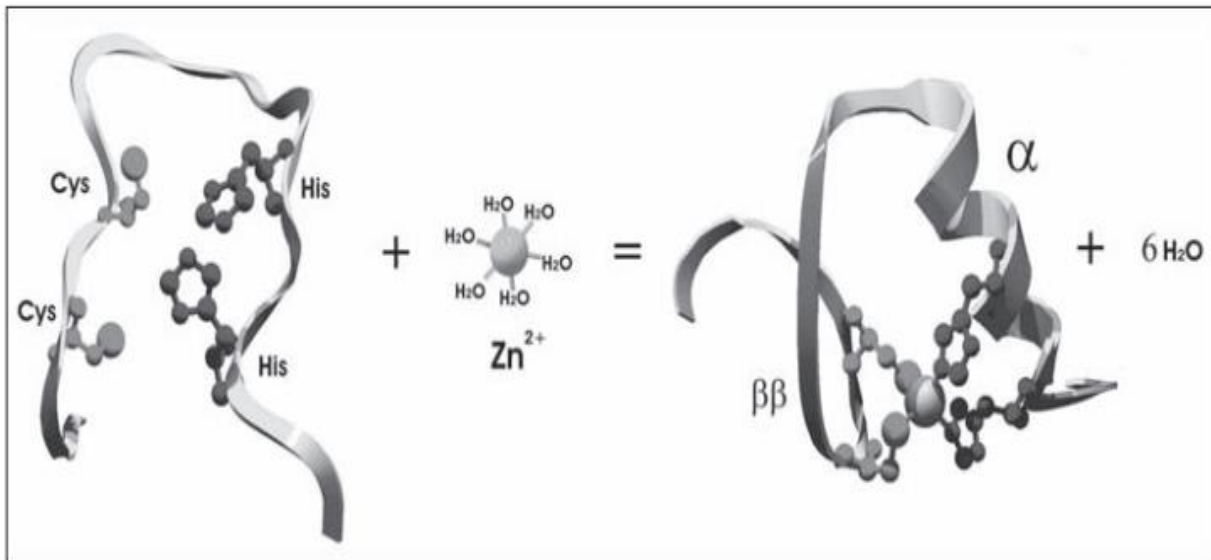


Figure 1.3: Zinc ions help to form DNA binding conformation of the protein<sup>8</sup>.

### **1.1.1.2 Tandem of C<sub>2</sub>H<sub>2</sub> Zinc Fingers**

C<sub>2</sub>H<sub>2</sub> zinc fingers have a tandem repeats of fingers connected by linkers "short oligopeptides". C<sub>2</sub>H<sub>2</sub> zinc fingers are classified to four classes depending on the number and the repeat pattern in the fingers: Single C<sub>2</sub>H<sub>2</sub>, triple C<sub>2</sub>H<sub>2</sub>, multiple adjacent C<sub>2</sub>H<sub>2</sub>, and the separated paired C<sub>2</sub>H<sub>2</sub> zinc finger proteins<sup>8</sup>.

A single zinc finger protein binds the DNA if there are additional domains to enhance its binding. Other proteins bind DNA without the requirement of other domains<sup>8,20</sup>.

The three zinc finger protein and multiple adjacent C<sub>2</sub>H<sub>2</sub> ZnFs bind the target DNA at three consecutive fingers<sup>5</sup>. The protein consisting of double fingers with a separation bind the target DNA specifically at one pair finger. This confirm the binding that most suitable unit to specifically bind the target DNA is the two to three C<sub>2</sub>H<sub>2</sub> zinc fingers. For example, TFIIA has nine zinc fingers that bind the DNA at fingers 1-3, it also touch the DNA at finger 5 and weakly at fingers 7-9<sup>21</sup>.

### **1.1.1.3 Features of the Binding Free Energy of C<sub>2</sub>H<sub>2</sub> Zinc Fingers to their Target DNA**

Gibbs free energy will change up on the binding of C<sub>2</sub>H<sub>2</sub> zinc fingers proteins to their target DNA by a value similar or higher than that for zinc finger folding (Table 1). This binding does not require any enzyme action<sup>8</sup>.

Table 1.1. Thermodynamics parameter for zinc finger reaction<sup>8</sup>.

Reaction	Number of Fingers	$k_a$ ( $10^8 M^{-1}$ )	$\Delta G$ (kcal/mol)	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)
$ZFP + Zn^{2+} \xrightleftharpoons{k_a} ZFP \cdot Zn^{2+}$	1	350	-8.8	-9.3	-0.5
$ZFP \cdot Zn^{2+} + DNA \xrightleftharpoons{k_a} ZFP \cdot Zn^{2+} \cdot DNA$					
Zif268	3	2.8	-11	-6.9	4.5
TFIIA	3	13	-9	-6.9	2.6
WT1	3	8.8	-12	7	19

As can be seen in Table 1.1,  $\Delta G$  values are always negative, so each of the above reactions prefer the association path over the dissociation reaction<sup>8</sup>.

#### 1.1.1.4 Three Zinc Finger Proteins versus Multiple Zinc Finger Proteins binding to DNA

The process through which zinc finger proteins recognize their target DNA is a big question. But the answer to this comes from the genetic oriented analysis of zinc finger-DNA interactions. For example, the structure of Zif268-DNA complex helps in the understanding of how triple fingers bind the DNA duplex<sup>22</sup>. The

binding of Zif268 through the three  $\alpha$  helices in the major groove of the target DNA antiparallel to the primary strand (the strand that zinc finger bind mostly) forming contacts to bases through hydrogen bonds, hydrophobic interactions and phosphate contacts<sup>8,22</sup>.

Zif268 bind the major groove of the DNA through the surface amino acid side chains at positions -1, 2, 3, 6 in its three  $\alpha$  helices and these side chains contact selectively with four sequential basses. Residues at positions -1, 3 and 6 bind to three basses of the primary strand and residue at position 2 binds to the fourth base in the complementary strand 5'-3'. Zif268 binds the ten residues base pairs "5'-GCGTGGGCGT-3'"<sup>8,12,22</sup>(Figure 1.4).

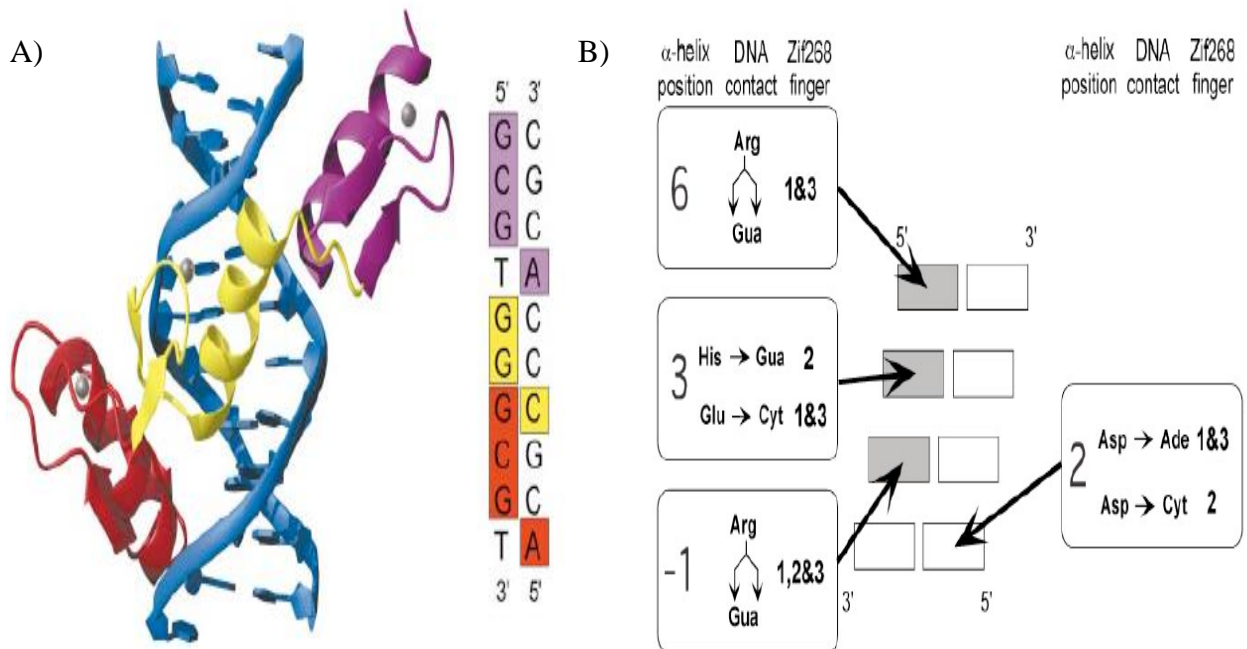


Figure 1.4: The Zif268-DNA complex. A) Major groove binding. Fingers 1, 2, and 3 of Zif268 are red, yellow and purple, DNA is blue. On the right is color coded parallel to finger color on the left. B) The side chain-base interactions in the Zif268 complex<sup>12</sup>.

Throughout the C<sub>2</sub>H<sub>2</sub> zinc fingers family the role of the side chains at positions -1, 2, 3, 6 is to make contact with bases, some exceptions were reported. For example, in the TFIIA-DNA complex there is an additional contact at position 10 and some side chains at the position -1, 2, 3, 6<sup>23</sup>.

Multiple zinc finger-DNA complexes exhibit irregular anomalous contacts due to the variation in the docking angle of the  $\alpha$  helix to DNA which expose the effect of all of the factors in specific zinc finger-DNA binding such as amino acid residues and adjacent zinc fingers<sup>8,12</sup>.

#### **1.1.1.5 Human PRDM9 allele-A ZnF Domain bound to DNA Sequence II<sup>24-28</sup>**

PRDM9 protein is involved in recombination process. In humans, the positions of recombination are chosen by a DNA-binding protein "PRDM9"<sup>24,25</sup>. PR domain having 9 (PRDM9) is a meiosis-specific histone H3 methyltransferase with the following functional domains: an N-terminal Kruppel-associated box "KRAB" domain; a central PR-SET domain and a C-terminal tandem array of a multiple C<sub>2</sub>H<sub>2</sub> ZnFs<sup>24,26</sup>(Figure 1.5).

C<sub>2</sub>H<sub>2</sub> zinc fingers are polymorphic within the same species these leads to variation in DNA-binding specificity. The PRDM9-DNA complex amino acid nucleotide binding does not follow the known binding model of C<sub>2</sub>H<sub>2</sub> zinc finger-DNA binding, i.e. -1,3,6<sup>26-28</sup>.

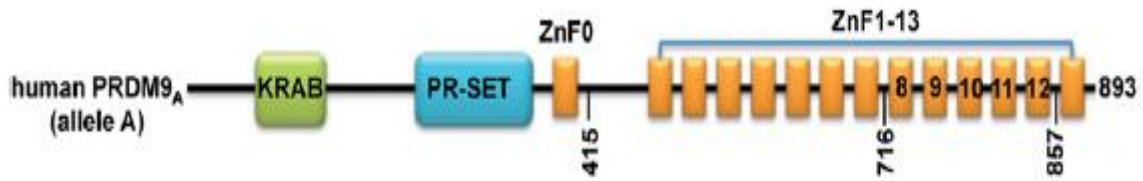


Figure 1.5: Domain organization of the hPRDM9 protein<sup>26</sup>.

The most common form of PRDM9 is allele A which is found in ~ 86% of European and ~ 50% of African populations<sup>28</sup>. The DNA binding specificity of zinc fingers 8-12 of hPRDM9 have a pattern of -NCCNCCNTNNCCNCN- which is involved in a ~ 40% of recombination hot spots<sup>26</sup> (Figure 1.6).

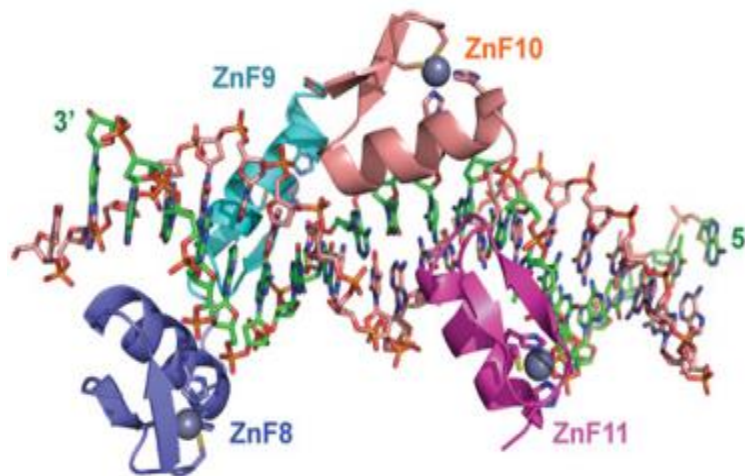


Figure 1.9: Crystal structure of hPRDM9A ZnF8–12 bound to THE1B sequence. ZnF8–11 (color-blue [ZnF8], cyan [ZnF9], orange [ZnF10], and magenta [ZnF11]) are shown in cartoon<sup>26</sup>.

### 1.1.2 Zinc Cluster Protein Zn<sub>2</sub>Cys<sub>6</sub> (Binuclear Cluster)

Cluster proteins with binuclear zinc ion were found in fungi "*Saccharomyces cerevisiae*" which contains a gene with 50 known zinc cluster proteins such as Gal4p<sup>29,30</sup>. Zn<sub>2</sub>Cys<sub>6</sub> Family of proteins have a conserved motif which composed of six cysteine residues " -Cys-X2-Cys-X6-Cys-X5-12-Cys-X2-Cys-X6-8-Cys- where X represents any amino acid<sup>31-33</sup> (Figure 1.7).

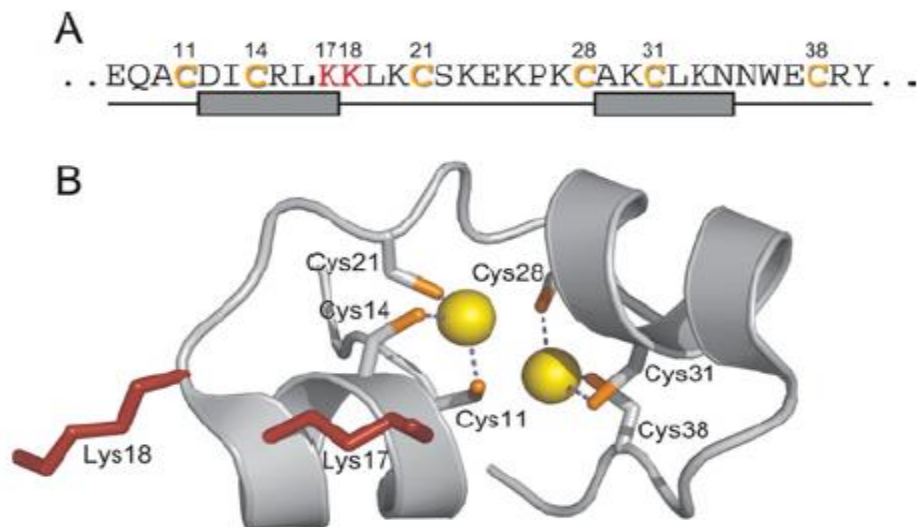


Figure 1.7: The Zn<sub>2</sub>Cys<sub>6</sub> zinc cluster motif. A) The amino acid sequence of the Gal4p Zn<sub>2</sub>Cys<sub>6</sub> zinc cluster motif. B) The binding of zinc ions facilitates the tight folding of the Zn<sub>2</sub>Cys<sub>6</sub> motif<sup>29</sup>.

#### 1.1.2.1 Structure - Function Relationship of Zinc Cluster Proteins

Zinc cluster proteins with different domains which function to bind DNA (Figure 1.8). A DNA binding domain which consist of three regions: the zinc finger; linker and dimerization regions. The regulatory region "MHR: middle homology region".



Finally the acidic region. These domains are responsible of DNA-binding specificity and protein-DNA interactions<sup>29</sup>.



Figure 1.8: Functional domains of zinc cluster<sup>29</sup>.

Zinc ion binding has two sites; each one is formed by three cysteine residues which are separated by a loop. They form together a pair of short  $\alpha$  helices that enfolded two zinc ions bridged by six cysteine residues<sup>29,34</sup>(Figure 1.9).

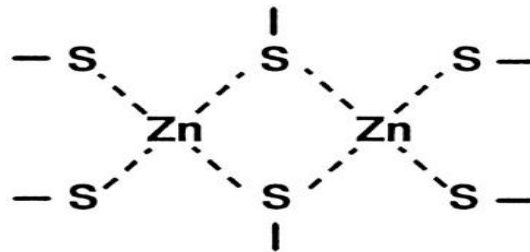


Figure 1.9: Zinc ion coordination in zinc cluster proteins<sup>35</sup>.

The linker region residues is the C-terminal of the zinc cluster region and it takes a different form in different zinc cluster proteins. This region contributes to the DNA-binding specificity because of its ability to recognize similar nucleotide on the preferred sequence and to prevent the binding to any other site. For example, Ppr1p, Gal4p, Put3p targets two CGG triplets<sup>29,31,35</sup> (Figure 1.10).

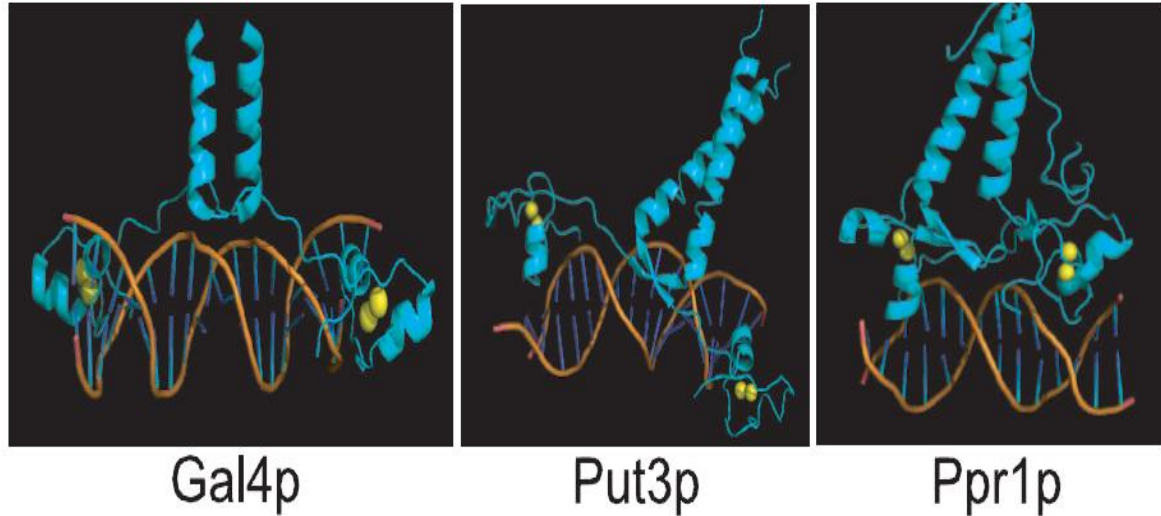


Figure 1.10: Crystal structures of the DBDs of Zn(II)<sub>2</sub>Cys<sub>6</sub> DNA binding proteins. zinc is represented as Yellow spheres<sup>29</sup>.

The dimerization region is present in the C-terminal region of the linker and it's found in most zinc cluster proteins. This region consists of repeats that form a coiled-coil structure and this mediates the homodimerization<sup>29,31</sup>.

The regulatory domain (middle homology region MHR) separates the DNA binding region from the acidic region. This region consists of about 80 amino acids, this region is important in the regulation of transcriptional activity of these proteins. The acidic domain is located on the C-terminal. Its function and structure varies and not well defined<sup>29,31,35,36</sup>.

### 1.1.2.2 Specificity of the DNA-Binding Process

$Zn_2Cys_6$  domain binds to the CGG sites in the major groove and to the phosphate backbone<sup>31</sup>. However, the spacing of the CGG triplets and their orientation are considered important factors in the DNA-binding specificity<sup>29</sup>.

These cluster proteins bind as homodimers to CGG triplets. These triplets adopt different orientation as are inverted, everted or direct repeats. Gal4p, Put3p and Ppr1p bind to inverted repeats. Leu3p and Pdr3p bind to everted repeats but Hap1p binding to a direct repeats because it has two zinc clusters in the same direction<sup>29,31</sup> (Figure 1.11).

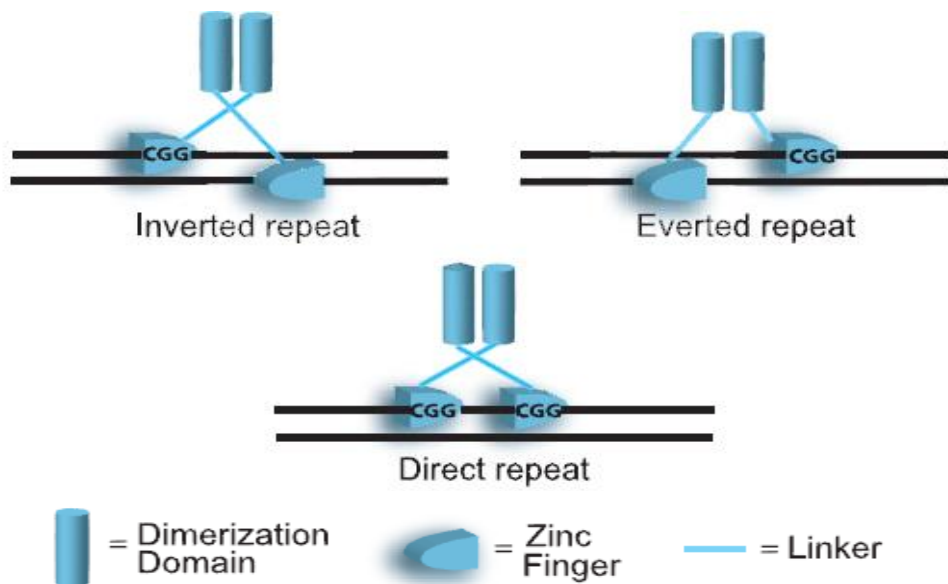


Figure 1.11: A model for zinc cluster protein DNA recognition<sup>31</sup>.



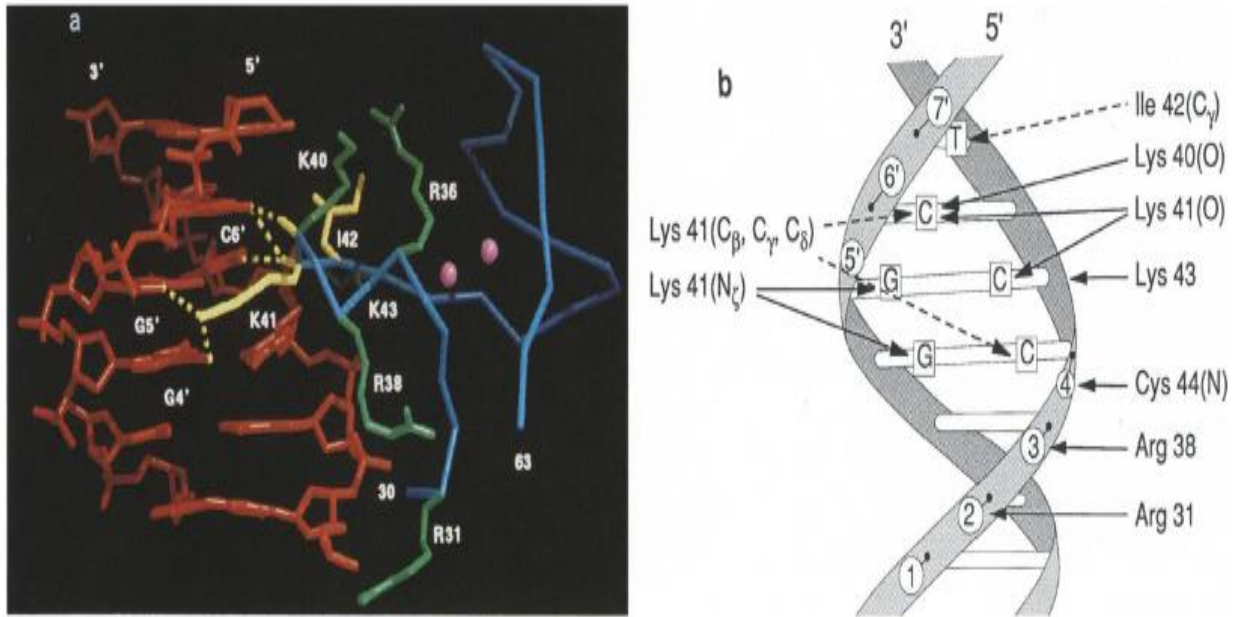


Figure 1.13: Zn domain bound to DNA. a) The interaction involves Zn domain. Protein is blue, the amino-acid is green, those interact with DNA phosphate oxygens are red. Residues with direct contact to DNA bases are yellow. Broken lines are hydrogen bonds between protein and DNA bases. b) phosphate and base contacts with protein. Van der Waals contacts are representing with broken lines<sup>38</sup>.

The linker in PPR1 folds into an anti-parallel  $\beta$ -sheet which has a 4 residue loop. In the two zinc domains the linkers have the same conformation but differ in their orientation with respect to the rest of molecule<sup>38</sup>.

#### 1.1.2.4 PUT3-DNA Complex

The *Saccharomyces cerevisia* PUT3 protein is a transcriptional activator of proline catabolic genes PUT1 and PUT2<sup>41-43</sup>. It contain a two Zn and six cysteine domain "Zn<sub>2</sub>Cys<sub>6</sub>" and it binds DNA on the two inverted CGG base pair which are separated by 10 base pair "CGG(n<sub>10</sub>)CCG"<sup>44</sup> (Figure 1.14).

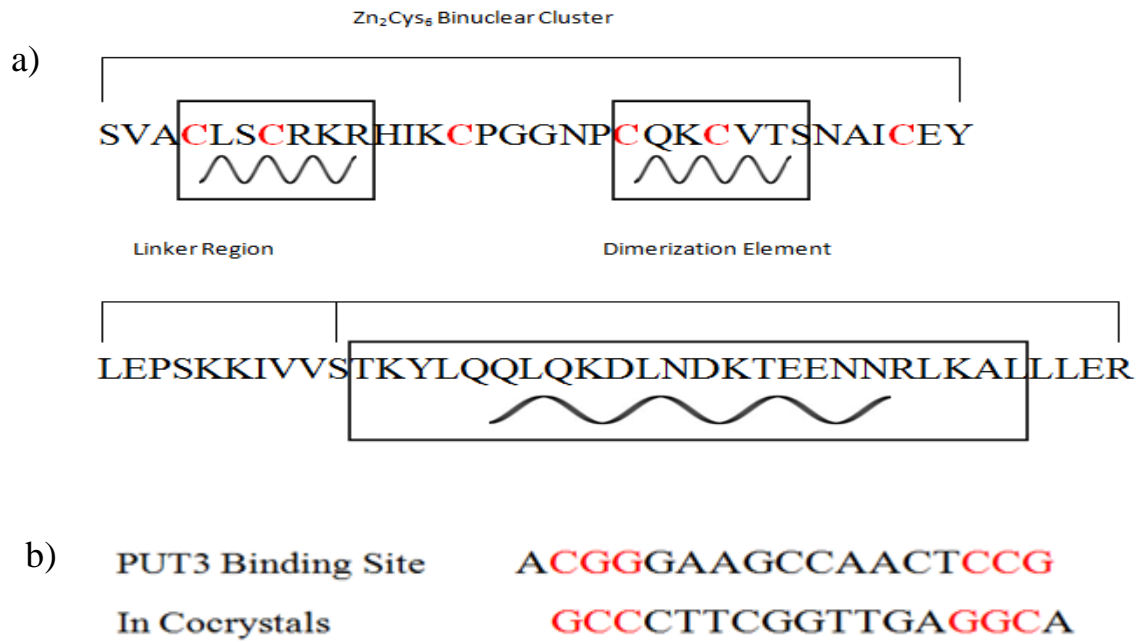


Figure 1.14: PUT3 DNA-binding domain and its DNA binding sites. a) Amino acid sequence of the PUT3 in the crystallographic analysis. b) 16-bp PUT3 binding site from the PUT2<sup>44</sup>.

PUT3 protein binds as nonsymmetrical homodimer to the DNA and the dimer twists around the DNA by one and a half turn (Figure 1.15). The connection between the two Zn domains and the two CGG triplets of the DNA are symmetrical and lie in the major groove at each end of the DNA sequence. The coiled-coil dimerization interface formed by the C-terminal end of the protein and it lie asymmetrically over the minor groove of the center, this helps to stabilize the dimer<sup>44</sup>.

The linkers of PUT3 are asymmetrically arranged and consist of  $\beta$  sheet regions lying between the minor groove of the DNA and the dimerization region. Van der

Waal and hydrogen bonds with the minor groove stabilize the asymmetry of the linkers<sup>44</sup>.

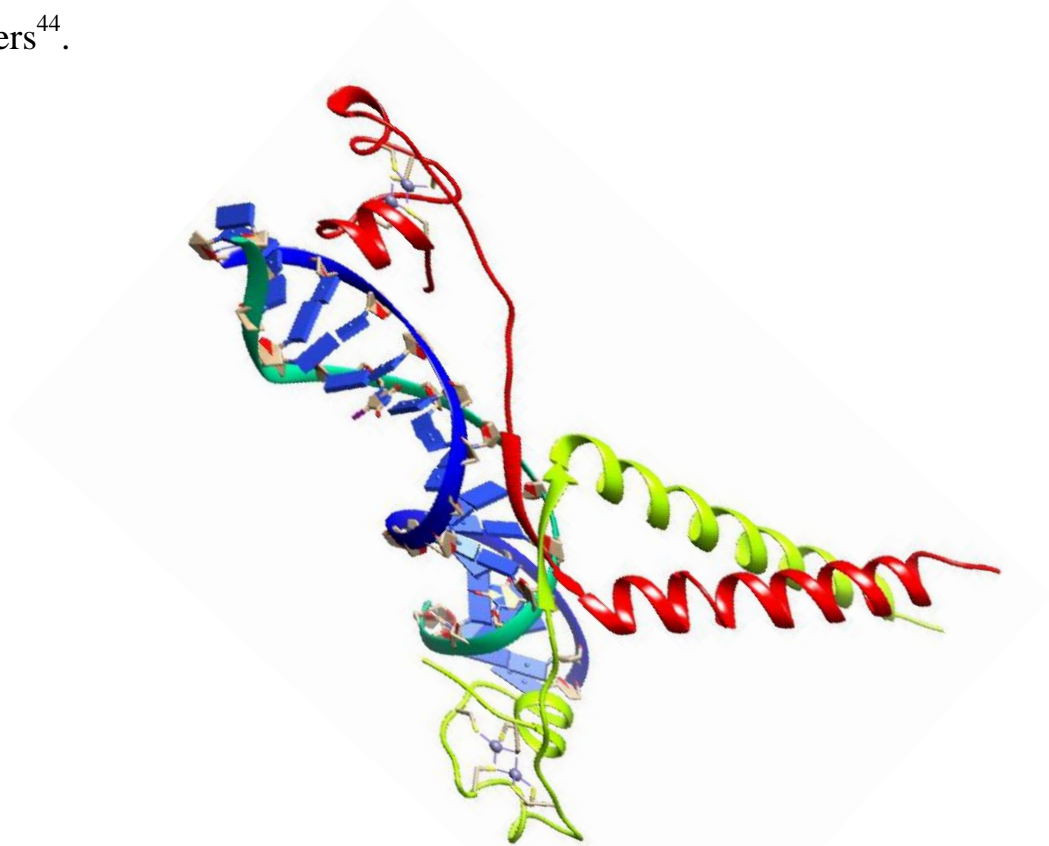


Figure 1.15: Overall structure of PUT3- DNA complex. DNA (red), CGG triplet (purple) metal ions (yellow)<sup>44</sup>.

The recognition of DNA by PUT3 consists of two parts. The first is recognition of CGG DNA sites by residues within the  $Zn_2Cys_6$  binuclear cluster, and the second is specificity of the 10 base pair sequence that separate the two inverted CGG by the linker and dimerization domain of the PUT3 protein<sup>44</sup>.

### 1.1.3 Biological Applications of Zinc Finger Proteins

ZnF proteins work as transcription factors in the human genome and play versatile roles in biological processes due to its varied combinations and functions<sup>45</sup>. They

are used as therapeutic agents to cure or prevent disease and they can also be used to control genes such as genes involved in cancer, cardiovascular disease, viral infection, and chronic pain<sup>46-49</sup>.

C<sub>2</sub>H<sub>2</sub> zinc finger proteins have important roles in cancer progression or suppression by regulating transcription of downstream genes such as proliferation, apoptosis, migration and invasion. Different levels of zinc finger protein regulation process in tumor genesis are summarized in Figure 1.16. First, ZnF proteins are regulated by cancer miRNA in different cancer types such as miR-199a-3p and miR-525-3p. Secondly, alarm cascades are activated by different environmental excitors then the ZnF worked through various PTMs such as phosphorylation and acetylation. PTMs regulation changes DNA binding abilities and affect it. Thirdly, ZnF uses different interacting proteins such as co-activators or co-repressors and other transcription factors in a way that ZnF can activate or suppress downstream genes. Finally, ZnF proteins give diverse specific DNA-binding abilities. The previous mechanism of ZnF in cancer progression differs according to type of cancer<sup>45</sup>.

Drugs that are specific to the expression or activation of C<sub>2</sub>H<sub>2</sub> ZnF proteins can be developed for a specific stage of cancer progression for therapeutic purposes<sup>45</sup>.



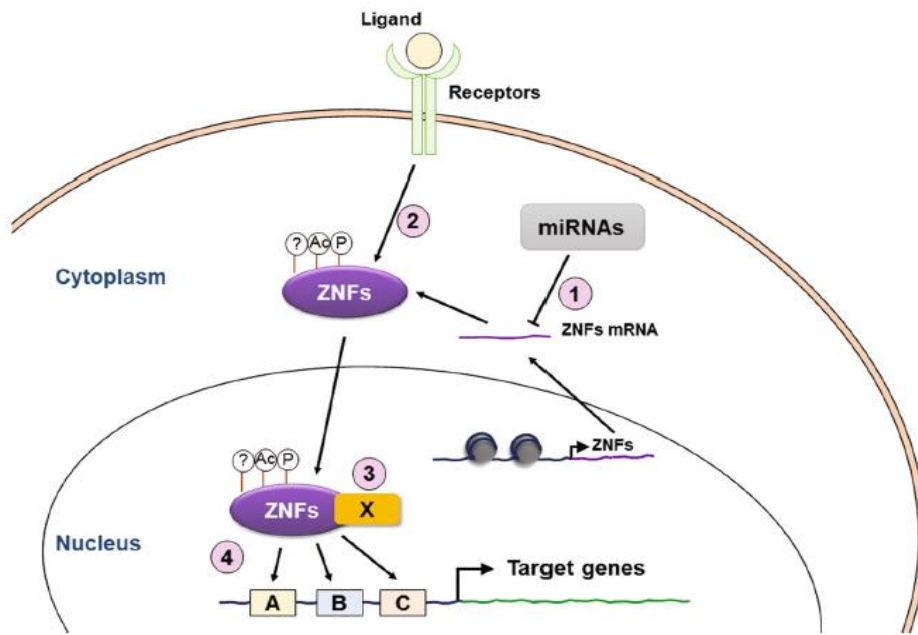


Figure 1.16: Different regulations of ZNF proteins functions in cancer progression<sup>45</sup>.

### 1.2.1 Binding Free Energy Calculations

Computational methods constitute some of the most important tools in the theoretical prediction of free energy of binding ( $\Delta G$ )<sup>50,51</sup>. Different computational methods are commonly used to predict  $\Delta G$ . In particular, Free energy perturbation (FEP) and thermodynamic integration (TI) are examples of methods that allow exact more accurate calculation of binding or solvation free energies but they are usually computationally expensive<sup>50-53</sup>. Both methods depends on simulations done at different intermediate stages between the unbound and bound end states of a protein and its target DNA<sup>51-53</sup>.

Other methods such as molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) and molecular mechanics/Generalized Born surface area (MM/GBSA) are less accurate but more computationally efficient<sup>54</sup>. They predict the binding free energy depending on the simulations of only bound and unbound end states of protein and its target DNA so they called end point, implicit solvent free energy methods<sup>52,53</sup>.

### 1.2.1 MM\_GBSA Method for Predicting the binding of Protein to DNA

Binding free energy of Protein to DNA complexation is calculated from the difference between the free energy of protein-DNA complex and the free energy of protein and the DNA separately as shown in Figure 1.17 and the following equation (1)<sup>53,55</sup>.

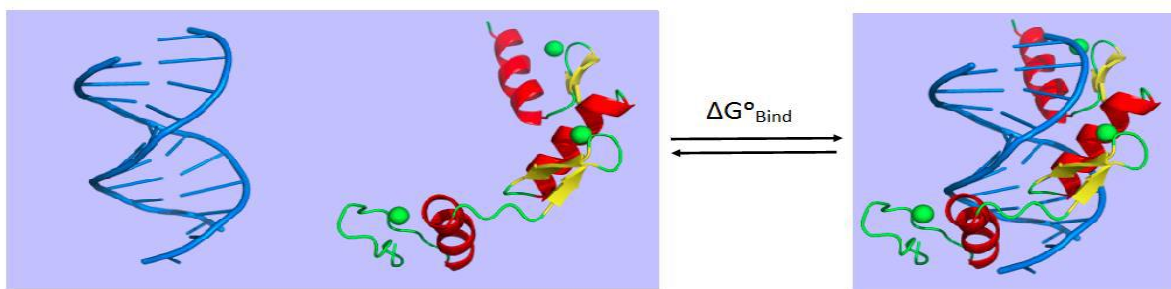


Figure 1.17: Binding free energy prediction for a solvated system as the difference between bound free energy and unbound free energy end states. Solvent shell is light blue<sup>55</sup>.

$$\Delta G^{\circ}_{\text{Bind}} = G^{\circ}_{\text{Protein-DNA complex}} - (G^{\circ}_{\text{Protein}} + G^{\circ}_{\text{DNA}}) \quad (1)$$

Most of the energy contribution comes from solvent-solvent interaction, and the fluctuation in total energy is larger than binding energy by ten times. Due to that the calculation time will be excessive even before the total energy converges to an acceptable tolerance. A more efficient method used to split the calculation as shown in the following thermodynamic cycle<sup>52,53,55,56</sup>(Figure1.18).

From the previous thermodynamic cycle the binding free energy can now be calculated from equation (2)<sup>53</sup>.

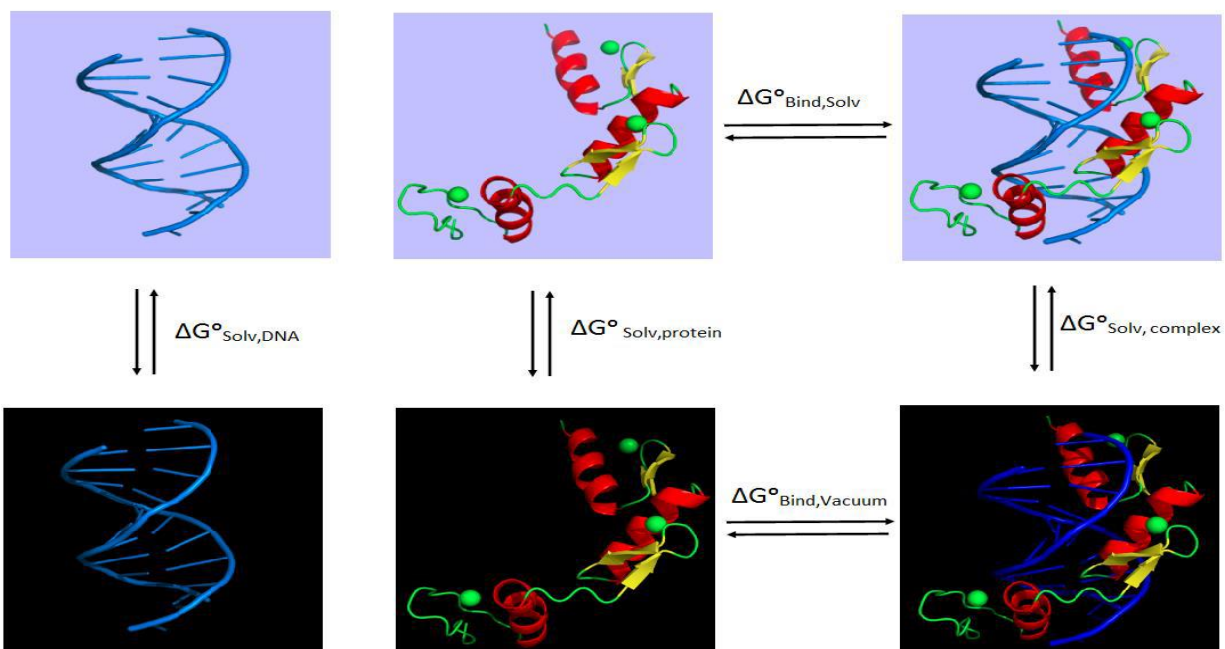


Figure 1.18: Thermodynamic cycle demonstrating steps of calculating binding free energy using MM/GBSA method. Solvent shell is light blue, vacuum is black<sup>55</sup>.

$$\Delta G^{\circ}_{\text{Bind, Solv}} = \Delta G^{\circ}_{\text{Bind, Vacuum}} + \Delta G^{\circ}_{\text{Solv, complex}} - (\Delta G^{\circ}_{\text{Solv, DNA}} + \Delta G^{\circ}_{\text{Solv, protein}}) \quad (2)$$

Solvation free energy can be calculated ( $\Delta G^{\circ}_{\text{solv}}$ ) by including polar and non-polar contribution as in equation 3.

$$\Delta G^{\circ}_{\text{Solv}} = \Delta G^{\circ}_{\text{Polar}} + \Delta G^{\circ}_{\text{non polar (Hydrophobic)}} \quad (3)$$

In the first step of solvation process, the non-polar contribution to free energy is proposed to be proportional to solvent surface area as in equation (4), where  $\gamma$  and  $\beta$  values depend on the applied solvation model and method<sup>53,54,57</sup>.

$$\Delta G^{\circ}_{\text{solv (non polar)}} = \gamma (\text{SASA}) + \beta \quad (4)$$

During the second step in the solvation process, the polar contribution is found from the difference in free energy that result from the transport of charged solute from gas phase with low dielectric constant ( $\epsilon=1$ ) to a solvent with high dielectric constant ( $\epsilon=80$ ) as in equation (5).

$$\Delta G^{\circ}_{\text{polar}} = G^{\circ}_{\text{electrostatic, } \epsilon=80} - G^{\circ}_{\text{electrostatic, } \epsilon=1} \quad (5)$$

In the Generalized Born model, each atom is represented as sphere with a radius  $\rho_i$  and charge  $q_i$ . A material of low dielectric constant ( $\epsilon=1$ ) fills the inner part of the sphere, and the atom present in a solvent with high dielectric constant ( $\epsilon=80$  for water at 300K) so the solvation free energy ( $\Delta G^{\circ}_{\text{polar}}$ ) in the GB methods can be calculated as in the following equation (6) and (7).

$$\Delta G_{\text{polar}} = - \sum_i (q_i^2 / 2R_i) (1 - 1/\epsilon_w) - \frac{1}{2} \sum_{ij, i \neq j} (q_i q_j / f^{\text{GB}}(r_{ij}, R_i, R_j))(1 - 1/\epsilon_w) \quad (6)$$

$$f^{\text{GB}} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{1/2} \quad (7)$$

Where  $\epsilon$  is the dielectric constant,  $q$  are the partial atomic charges,  $r_{ij}$  the inter-atomic distances, and  $R_i$  are the born radii calculated according to the pairwise descreening algorithm of Hawkins et al<sup>58,59</sup>.

On the other hand, the solvation energies ( $G_{\text{polar}}$ ) for PB are obtained by using a grid based finite difference solution to the Poisson Boltzmann equation with zero salt concentration (equation 8), where  $\rho(r)$  is the charge distribution of the molecule,  $\epsilon(r)$  is the dielectric constant,  $\Phi(r)$  is the electrostatic potential.  $G_{\text{polar}}$  for PB is calculated as a sum over all atoms in a system described by classical force fields which yields the electrostatic potential at every grid point (equation 9), where the partial atomic charge ( $q$ ) for atom  $i$  is multiplied by the difference in the calculated grid-point potential  $\Phi_i$  for the transfer from gas phase ( $\epsilon=1$ ) to the water ( $\epsilon = 80$ )<sup>58,59</sup>.

$$\nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \quad (8)$$

$$G_{\text{polar}} = 1/2 \sum_i^N q_i(\phi_i^{80} - \phi_i^1) \quad (9)$$

The binding free energy in vacuum ( $\Delta G^{\circ}_{\text{vacuum}}$ ) can be predicted by calculating the average interaction energy between the protein and DNA in gas phase ( $\Delta E^{\circ}_{\text{Molecular mechanics}}$ ) and the entropy change due to binding as shown in equation (10).

$$\Delta G^{\circ}_{\text{Bind, Vacuum}} = \Delta E^{\circ}_{\text{Molecular mechanics}} - T \cdot \Delta S^{\circ}_{\text{Normal mode analysis}} \quad (10)$$

The average interaction energies between protein and DNA are obtained by using molecular mechanics (MD) simulations by applying force field functions and parameters which are composed of two types of energy. First, the covalent energy represented by bonds, angles and dihedral energies. Second, the non-covalent energies represented by electrostatic and van der Waals energies as in equation (11).

$$\mathbf{E}_{\text{MM}} = \mathbf{E}_{\text{Bond}} + \mathbf{E}_{\text{Angle}} + \mathbf{E}_{\text{Torsion}} + \mathbf{E}_{\text{van der Waals}} + \mathbf{E}_{\text{Electrostatic}} \quad (11)$$

The entropy change can be calculated using normal mode analysis (*Nmode*). This term is usually ignored when calculating relative binding free energy because due to the fact that calculating entropy is a demanding process that adds a little information when similar ligands are bound to the protein<sup>52,54,60,61</sup>.

## Chapter 2

### COMPUTATIONAL METHODS

#### 2.1 Initial Structures of Protein-DNA Complexes

The structures of three zinc finger-DNA complexes were obtained from the Brookhaven Protein Data Bank (RSCB). PRDM9 allele-A ZnF Domain with bound to DNA Sequence (PDB code: 5EGB). PPR1-DNA Complex (PDB code: 1PYI) and PUT3-DNA Complex (PDB code: 1ZME). All crystal water molecules were removed.

5EGB zinc finger-DNA complex contains four zinc fingers. Each zinc ion tetrahedrally coordinated to two histidine and two cysteine residues in the deprotonated form.

PPR1 and PUT3 contains four zinc ions in a dimer form ( $Zn_2Cys_6$ ). Each Zn ion is bonded to 4 cysteine residues; two of the cysteine residues are shared between two Zn ions as shown Figure 2.1<sup>32,38,44</sup>.



Figure 2.1: Zinc ion coordination in zinc cluster proteins<sup>35</sup>.

The *Prmtop* and *Inpcrd* Files for AMBER were created using *Xleap* editor, and different mutations to protein-DNA complexes were prepared using *PyMOL* and *USCF chimera* programs (Appendix A).

## **2.2 Equilibration of the solvated systems**

The three systems under study were equilibrated in four steps before performing molecular dynamic production stage: minimization, heating, density equilibration, and unrestrained equilibration.

### **a- Minimizing the Solvated Systems**

Minimization releases the system from the stress which is formed by unfavorable van der Waal and electrostatic interactions. When our systems were solvated the pre-equilibrated TIP3P water surrounded the solute and forms some gaps between the solvent and solute. After the minimization, water molecules become more relaxed with the solute forming a more stable solvated structure.

500 steps of the steepest descent method followed by 500 steps using the conjugate gradient algorithm at constant volume periodic boundary condition were performed to minimize the system. The *mdin* file that used to perform minimization was titled *min.in* (Appendix D).



## **b- Heating the Solvated Systems**

Gradual heating of the solvated system to 300K using *Langevin* dynamics were performed. The heavy atoms of complexes were restrained by a harmonic potential with a force constant of 2 kcal/mol-Å<sup>2</sup>. The input file was *heat.in* which was used to perform heating (see Appendix D).

## **c- Density Equilibration of the Solvated Systems**

50 picosecond of density equilibration was performed at 300K with constant pressure periodic boundary conditions and a harmonic potential positional restrains of 2 Kcal/mol-Å<sup>2</sup>. File used *density.in* (Appendix D).

## **d- Unstrained Equilibration of the Solvated Systems**

Further equilibration was applied to the system for a 500 picoseconds of unstrained equilibration at 300K and constant pressure. The SHAKE method was used to hold all hydrogen-heavy atom bond distances. The title of the file used in this step was *equil.in* (Appendix D).

## **2.3 Analysis of Output Files from Equilibration Process**

Over the simulation process, the studied systems are supposed to reach and maintain an equilibrated state. This was checked by observing the changes in different properties of the systems during the simulation process<sup>62</sup>. The properties

of systems were extracted from the output files of various equilibration runs. The plotted density, temperature, pressure and energy versus simulation time of the 5EGB system are shown in Figure 2.2. Other plots for PPR1 and PUT3 complexes can be seen in Appendix E.

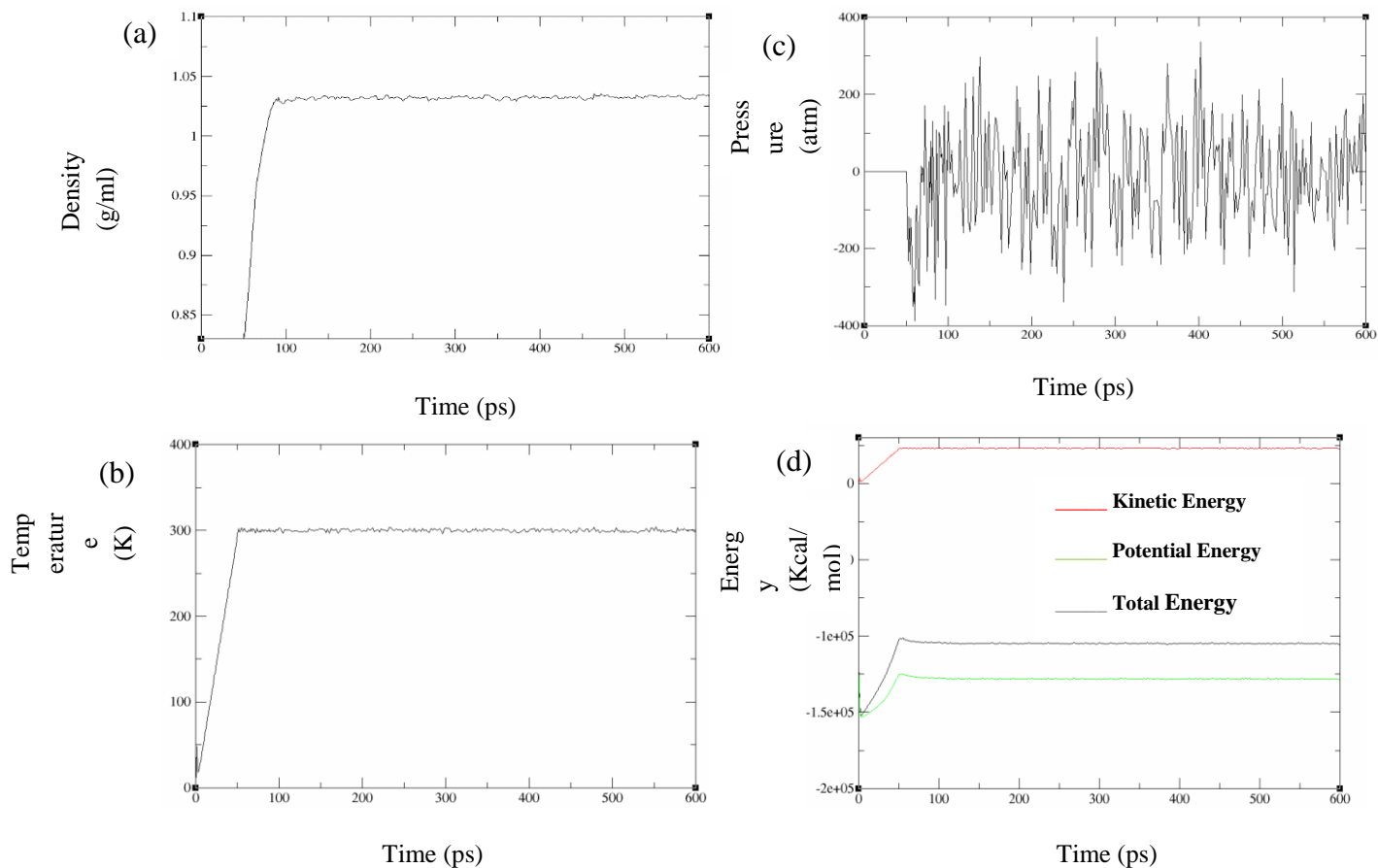


Figure 2.10: Plots against time for the heating and equilibration phases of the density (a), temperature (b), pressure (c), and energy (d).

Density values are not written to the output file in the first 50ps because the heating stage was carried out under constant volume. The system was equilibrated to a density of approximately 1.032 g/ml (the density of pure water at 300K 1g/ml)

(Figure 2.2 (a)).The added protein-DNA complex lead to slight increase in the density of the system<sup>63</sup>.

Figure 2.2 (b) shows the change in system temperature as a function of simulation time, started at 10 K and rises to 300 K over a period of about 50ps. The temperature after 50ps remained at 300K indicating that the system uses langevin dynamics effectively<sup>63</sup>.

Figure 2.2 (c) shows the change in the system pressures as a function of simulation time. It started at zero during the 50ps heating step because the system was under constant volume. Pressure values decreased to negative over a period 50-80ps because of a force that worked to reduce the volume of the water box and the positive values due to force worked to increase it. The pressure after 80ps remained at around 1atm<sup>63</sup>.

Figure 2.2 (d) shows the change in the system energies as a function of simulation time. Energies were increased during the first 50ps due to the heating from 10 to 300K. The kinetic energy after 50ps stayed constant indicating a successful temperature thermostat. The potential energy plot shows slight decreases indicating that the first system relaxation under constant pressure was successful and for the rest of simulation it becomes stable. Total energy plot was consistent with the behavior of kinetic and potential energy<sup>63</sup>.

## 2.4 Production Simulations of the Solvated Systems

Two nanoseconds of production runs were performed at the same conditions of the previous equilibration steps to prevent any sudden jump in potential energy. Molecular dynamic snapshots trajectories were collected every 10 picoseconds and production was run 4 times to obtain 2ns simulation time. The title of our used file is *prod.in* (Appendix D).

## 2.5 Calculating the Binding Free Energy of Protein-DNA complexes

Binding free energy of 5EGB, PPR1 and PUT3 bound to their specific DNA bases were calculated using MM/GB(PB)SA methods which involves the following steps:

a- The dry proteins PDB files were split into two files: the DNA file and the Zinc finger protein file.

b- Binding free energy of the complexes was calculated using the input file **calculate-energy.in** (see Appendix D). The PDB files needed were: solvated complex, DNA and zinc finger PDB files and the coordinate files from the production run.

## 2.6 Calculation of the Entropy Contribution

Normal mode analysis (*Nmode*) was used to calculate the entropy contribution by calculating the normal mode for DNA and zinc finger protein complex. The results were averaged to obtain an estimation of the binding entropy. The input file that was used for calculation is **calculate-entropy.in** (Appendix D).

## 2.7 Analysis of Results

The temperature versus time and total kinetic energy versus time and other summary output files were plotted using *xmgrace* program. Root mean square deviation (RMSD) for DNA and protein backbone was calculated using *cpptraj*, input files *measure\_DNA\_rmsd.ptraj* and *measure\_protein\_rmsd.ptraj*.

Bond length and angle for hydrogen bonds between DNA and the proteins amino acid side chains were calculated using *cpptraj* and *analyse\_hbond.in* input file. The production files were converted to *binpos* file using *mdcrd\_to\_binpos.ptraj* input file. The length of hydrogen bond is the distance between donor and acceptor atom, and the bond length was considered if in the range of 2.2-2.5 Å as strong bond, considered as moderate bond if its length is in the range 2.5-3.2 Å and a weak bond if the length is 3.2-4.0 Å<sup>64</sup>.

# Chapter 3

## RESULTS AND DISCUSSION

### 3.1 Protein-DNA Interaction

#### 3.1.1 Binding of PRDM9<sub>A</sub> and Its Alleles to DNA

From the MD simulations, the RMSD for the complex backbone (CA, C, N) has been plotted (Figure 3.1). It shows that RMSD values fluctuated steadily around 2 Å for PRDM9<sub>A</sub>-DNA complex indicating reasonable stability upon simulation<sup>65</sup>.

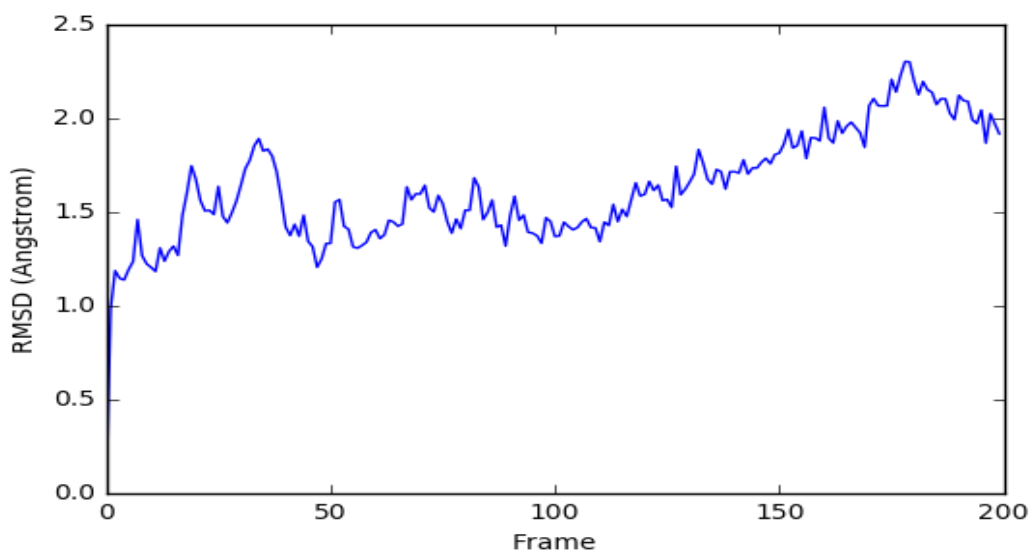
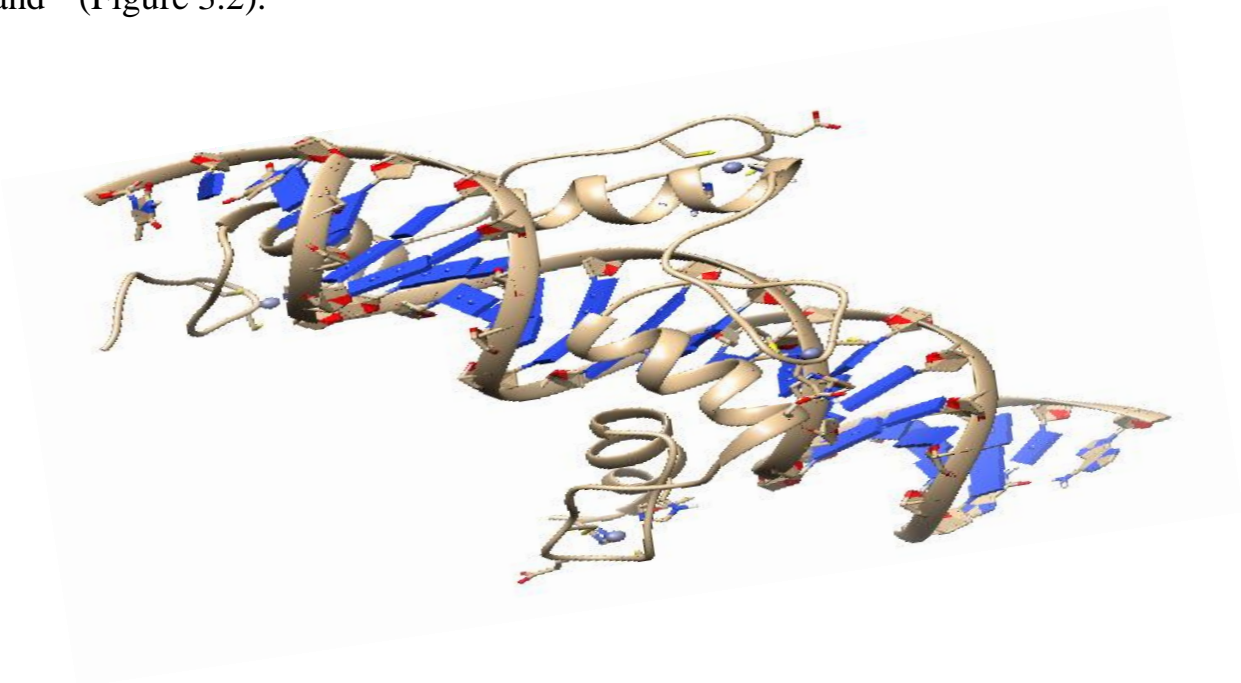


Figure 3.1 : Backbone (CA, C, N) RMSD vs. frame (reference residue number) after the production phase of the MD refinement for PRDM<sub>A</sub>-DNA complex.

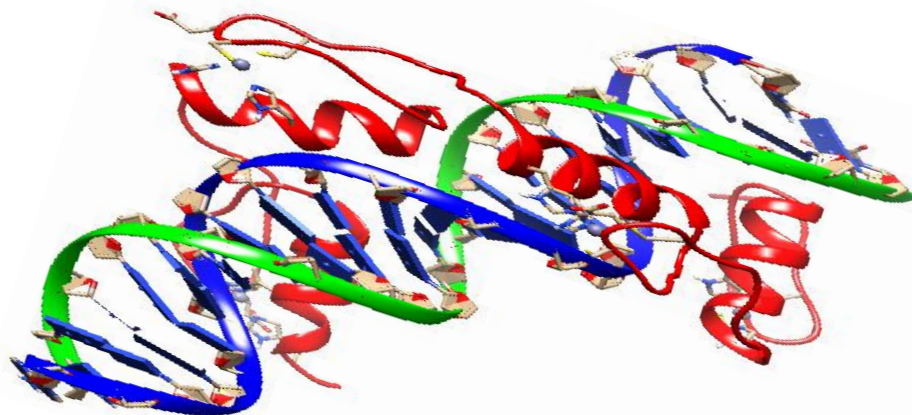
### 3.1.1.1 Overall Structure of PRDM9<sub>A</sub>-DNA Complex

The average structure of the complex PRDM9<sub>A</sub>-DNA from molecular modeling simulations is shown in Figure 3.2. Like other conventional C<sub>2</sub>H<sub>2</sub> ZnF proteins, each zinc finger in PRDM9<sub>A</sub> contains two  $\beta$ -strand and an  $\alpha$  helix through which it contacts three to four DNA base pairs. After its binding to DNA each  $\alpha$  helix lies in the major groove of the DNA. Also the  $\beta$ -strands and the Cys<sub>2</sub>-Zn-His<sub>2</sub> lie on the outside. Each  $\alpha$ -helix makes hydrogen bonding between three to four base nucleotides, most of these nucleotides are purines of the complementary DNA strand<sup>26</sup> (Figure 3.2).

A)



B)



C)

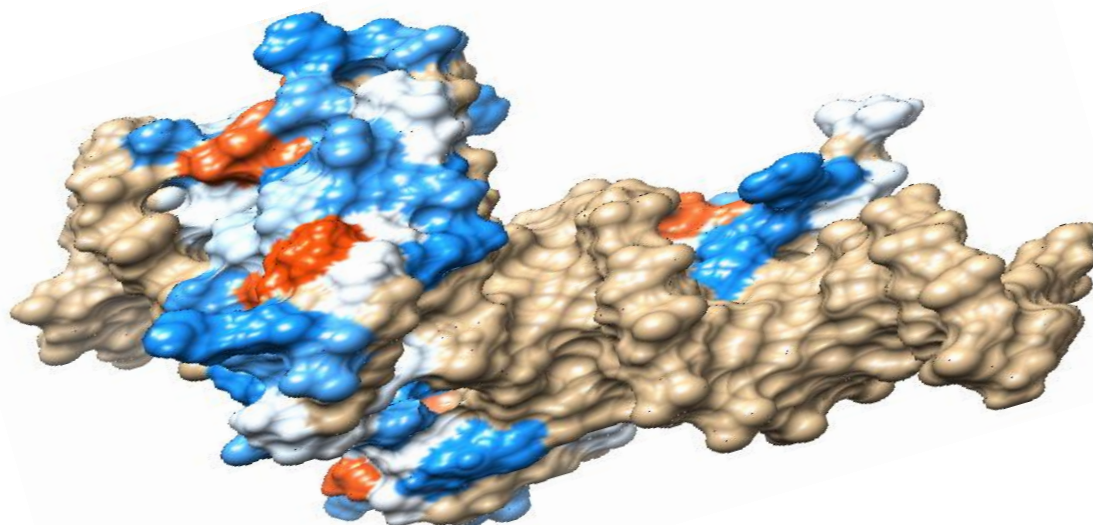


Figure 3.2 : Overall structure of PRDM9<sub>A</sub>-DNA complex after molecular modeling simulation, showing the four zinc fingers. A) ZnF1-4 in complex with the DNA strands , ZnF1-4 are represented as a light brown and the coordinated zinc ions are represented as blue spheres. B)The PRDM9<sub>A</sub> complex are represented as a rounded ribbon, ZnF1-4 are red and the coordinated zinc ions are represented as a blue spheres. C) The PRDM<sub>9</sub> complex are represented as a hydrophobisity surface.



### 3.1.1.2 Binding Affinities of PRDM9 Alleles to DNA

Most of the binding energies have a negative value indicating a high affinity of ZnF proteins to the DNA sequences studied. The complexes also exhibit an enthalpy driven process indicating a major groove binding and resulting in no considerable DNA deformation<sup>66,67</sup>(Table 3.1 and Figure 3.3)..

Table 3.1. Calculated energies (kcal/mol) for 1AAY, and PRDM<sub>9</sub>-DNA with different alleles complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM-GBSA and MM-PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting  $T\Delta S$  from *nmode* calculations from the binding free energy.

Complexes	$\Delta G^{\text{binding}}$ ( $\Delta H$ ) (MM/GBSA) kcal/mol	$\Delta G^{\text{binding}}$ ( $\Delta H$ ) (MM/PBSA) kcal/mol	$T\Delta S$ Kcal/mol	$\Delta G^{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$	$\Delta G^{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$
<b>1AAY</b> <b>(3F)</b>	-84.09 $\pm$ (4.03)	-106.7 $\pm$ (6.31)	-88.36 $\pm$ (5.78)	+4.270 $\pm$ (4.03)	-18.38 $\pm$ (6.31)
<b>5EGB</b> <b>(4F)</b>	-145.3 $\pm$ (5.65)	-190.5 $\pm$ (7.29)	-115.8 $\pm$ (11.82)	-29.48 $\pm$ (5.65)	-74.66 $\pm$ (7.29)
<b>L9/24</b> <b>(4F)</b>	-145.8 $\pm$ (8.82)	-187.7 $\pm$ (10.39)	-113.9 $\pm$ (7.30)	-31.92 $\pm$ (8.82)	-73.77 $\pm$ (10.39)
<b>L13</b> <b>(4F)</b>	-155.8 $\pm$ (7.77)	-219.1 $\pm$ (11.16)	-127.9 $\pm$ (4.94)	-27.83 $\pm$ (7.77)	-91.19 $\pm$ (11.16)
<b>L20<sub>bindTA</sub></b> <b>(4F)</b>	-141.5 $\pm$ (9.23)	-198.4 $\pm$ (11.390)	-119.1 $\pm$ (7.87)	-22.44 $\pm$ (9.23)	-79.33 $\pm$ (11.36)
<b>L20<sub>bindCG</sub></b> <b>(4F)</b>	-133.6 $\pm$ (7.64)	-177.8 $\pm$ (9.92)	-123.6 $\pm$ (4.04)	-9.970 $\pm$ (7.64)	-54.20 $\pm$ (9.92)

Table 3.2. Calculated energies for 1AAY, and the variant alleles of PRDM9-DNA complex using MM/GBSA. The Van der Waals energy (VdW), electrostatic energy ( $G_{ELE}$ ), the polar energy ( $E_{GB}$ ), and the non-polar energy ( $E_{SURF}$ ) are shown.

<b>Complexes</b>	<b>VdW</b> kcal/mol	<b>STD</b> kcal/mol	<b><math>G_{ELE}</math></b> kcal/mol	<b>STD</b> kcal/mol	<b><math>E_{GB}</math></b> kcal/mol	<b>STD</b> kcal/mol	<b><math>E_{SURF}</math></b> kcal/mol	<b>STD</b> kcal/mol
<b>1AAY</b> <b>(3F)</b>	-110.1	5.14	-5191	18.43	5237	16.73	-19.52	0.20
<b>5EGB</b> <b>(4F)</b>	-175.4	7.61	-1116	29.48	11210	27.70	-30.29	0.22
<b>L9/24</b> <b>(4F)</b>	-173.8	7.52	-9840	24.44	9898	23.51	-30.23	0.23
<b>L13</b> <b>(4F)</b>	-175.4	8.99	-11871	25.09	11871	24.03	-30.50	0.20
<b>L20<sub>bindTA</sub></b> <b>(4F)</b>	-172.8	9.08	-11140	33.33	11201	31.17	-30.22	0.17
<b>L20<sub>bindCG</sub></b> <b>(4F)</b>	-172.3	8.18	-11185	25.31	11254	23.70	-30.14	0.21

Table 3.3. Calculated energies for 1AAY, and the variant alleles of PRDM9-DNA complex using MM/PBSA. The Van der Waals energy (VdW), electrostatic energy ( $G_{ELE}$ ), the polar energy ( $E_{PB}$ ), and the non-polar energy ( $E_{nopolar}$ ) are shown.

<b>Complexes</b>	<b>VdW</b> kcal/mol	<b>STD</b> kcal/mol	<b><math>G_{ELE}</math></b> kcal/mol	<b>STD</b> kcal/mol	<b><math>E_{PB}</math></b> kcal/mol	<b>STD</b> kcal/mol	<b><math>E_{nopolar}</math></b> kcal/mol	<b>STD</b> kcal/mol
<b>1AAY</b> <b>(3F)</b>	-110.1	5.14	-5191	18.43	5211	16.89	-15.93	0.10
<b>5EGB</b> <b>(4F)</b>	-175.4	7.61	-1116	29.48	11157	27.82	-22.27	0.13
<b>L9/24</b> <b>(4F)</b>	-173.8	7.52	-9840	24.44	9848	23.92	-22.06	0.11
<b>L13</b> <b>(4F)</b>	-175.4	8.99	-11871	25.09	11799	24.18	-22.24	0.10
<b>L20<sub>bindTA</sub></b> <b>(4F)</b>	-172.8	9.08	-11140	33.33	11136	32.10	-22.30	0.10
<b>L20<sub>bindCG</sub></b> <b>(4F)</b>	-172.3	8.18	-11185	25.31	1120	24.40	-22.46	0.11

Figures 3.3- 3.5 show different binding free energies for the variant alleles of PRDM9 protein. Despite the difference in energy values as shown in Figure 3.3, the trend in energy variation is almost the same with a slight different behavior for different alleles in both MM/GBSA and MM/PBSA. The entropy change is negligible and this confirms that the binding energy ( $\Delta H$ ) is the driving force in

this process<sup>8,66,67</sup>. The changes in values of absolute energy  $\Delta G$  are parallel to the change in the binding energy ( $\Delta H$ ) in both MM/GBSA and MM/PBSA. Binding energy ( $\Delta H$ ) correlates better with non-polar energy ( $E_{\text{SURF}}$ ) ( $R=0.97$ ), and Van der Waals ( $R = 0.97$ ) in MM/GBSA method than the electrostatic ( $E_{\text{ELE}}$ ) ( $R=0.94$ ), and polar ( $E_{\text{GB}}$ ) ( $R=0.94$ ) energies. But in MM/PBSA the binding energy ( $\Delta H$ ) correlates with electrostatic ( $E_{\text{ELE}}$ ) ( $R=0.95$ ), and polar ( $E_{\text{PB}}$ ) energies ( $0.95$ ). The correlation with non-polar energy and Van der Waals are slightly lower ( $R=0.93$ , and  $R=0.94$ , respectively). In MM/GBSA the contribution for both  $G_{\text{ELE}}$  and  $E_{\text{GB}}$  and this made the energy outcome correlates better with experimental values<sup>68</sup> (see Figure 3.4, 3.5 and Appendix F).

The difference in the energy outcome of these two methods (MM/PBSA & MM/GBSA) is mainly due to the different calculation approach for each one. The *Generalized Born* (GB) model approximates the exact Poisson-Boltzmann equation and treats the solute as a set of spheres whose internal dielectric constant differs from the external solvent (water  $\epsilon=80$ ), while in Poisson-Boltzmann equation (PB) has less approximations and describes the electrostatic environment of a solute in a solvent containing ions<sup>58</sup> (see section 1.2.1, and 3.2).

The good correlation with the total electrostatic energy (the sum of the electrostatic energy as calculated by molecular mechanics (ELE) and the electrostatic contribution to the solvation free energy calculated by GB/PB methods in both

methods) indicates a good role for direct and water-mediated hydrogen bonding in protein-DNA interaction in all cases<sup>68</sup>.

Table 3.4: Correlation coefficients (R) for different energy contributions to the total free energy of binding ( $\Delta H$ ) in both MM/GBSA and MM/PBSA methods

Energy Component	Correlation Coefficient (R)	Correlation Coefficient (R)
	(MM/GBSA)	(MM/PBSA)
<b>GELE</b>	0.9442	0.9503
<b>VdW</b>	0.9706	0.9427
<b>EGB/PB</b>	0.9432	0.9512
<b>ESURF/npolar</b>	0.9670	0.9269

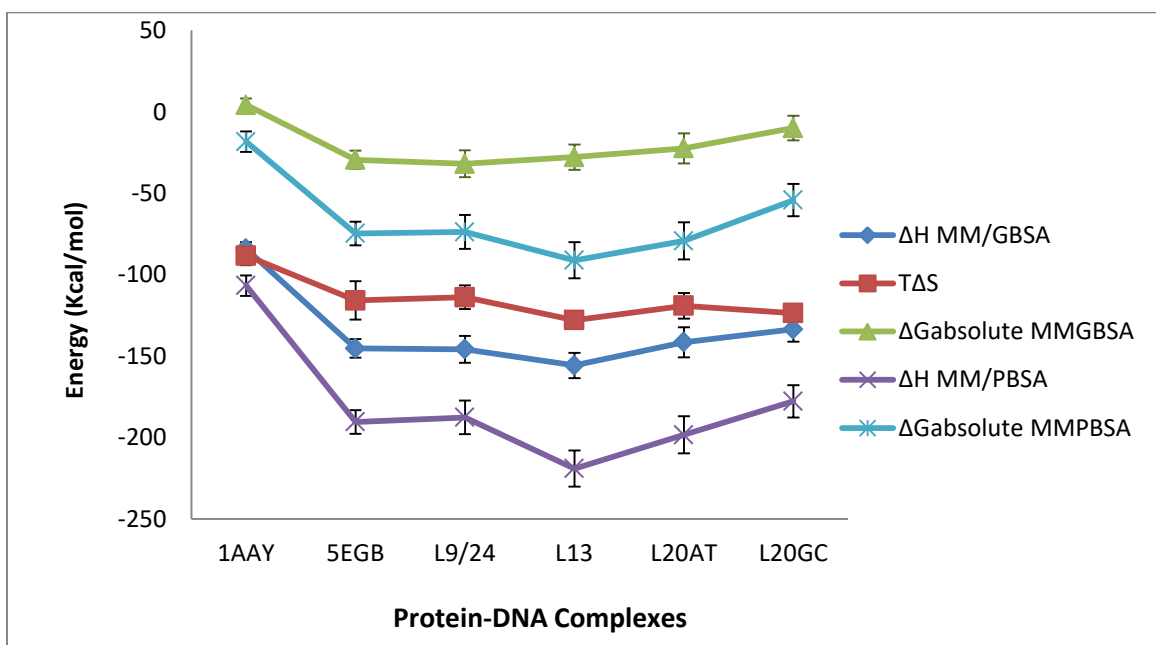


Figure 3.3: The change in the absolute energy ( $\Delta G_{\text{absolute}}$ ) for different alleles of PARMD9. Binding energy ( $\Delta H$ ) and entropy contribution ( $T\Delta S$ ) to the absolute energy from both MM/GBSA and MM/PBSA. Absolute energy change for  $\Delta G_{\text{MM/GBSA}}$  labeled green and for  $\Delta G_{\text{MM/PBSA}}$  orange for different alleles. Binding energy  $\Delta H_{\text{MM/GBSA}}$  labeled blue and  $\Delta H_{\text{MM/PBSA}}$  purple). Entropy contribution ( $T\Delta S$ ) labeled red.

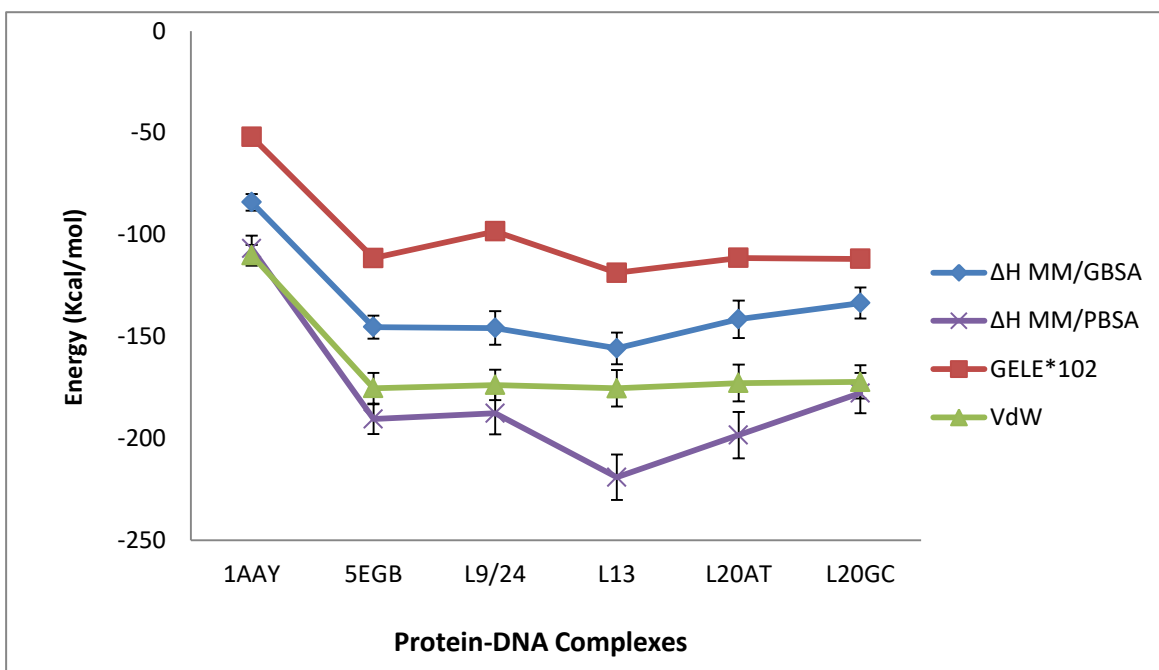
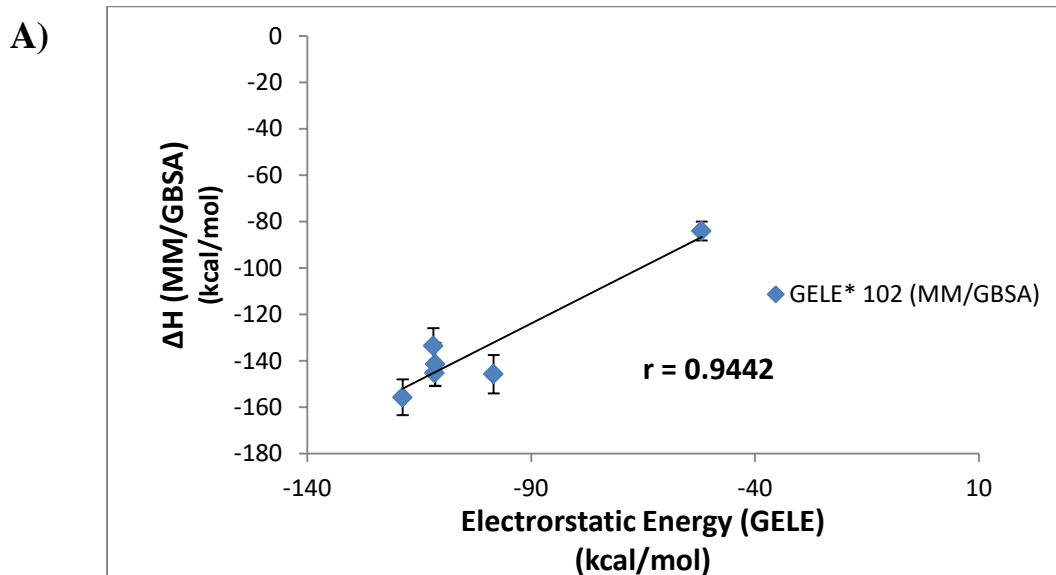


Figure 3.4: Change in the binding energy for different alleles of PARMD9. Binding energy ( $\Delta H_{MM/GBSA}$  and  $\Delta H_{MM/PBSA}$ ) are represented as blue and purple, respectively.  $G_{ELE}$  is represented as a red squares with red line, VdW energy is represented as a green triangles with green line. Electrostatic energy multiplied by  $10^2$ .



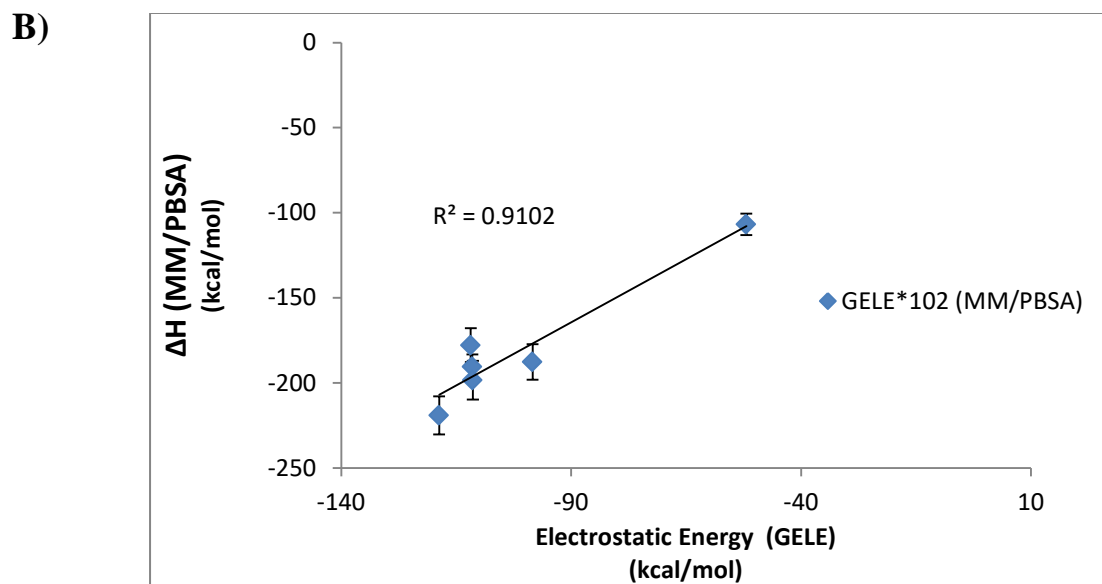


Figure 3.5: Correlation between binding energy ( $\Delta H_{MM/GBSA}$  and  $\Delta H_{MM/PBSA}$ ) and electrostatic energy ( $G_{ELE}$ ) of amino acid residues upon different alleles of PARMD9 using in A) MM/GBSA methods and in B) MM/PBSA method. The  $r$  values of 0.94 and 0.95 indicates a strong correlation between the binding energy and electrostatic energy.

### 3.1.1.3 Binding of PRDM9<sub>A</sub> Compared to the three Finger protein 1AAY to DNA

Comparing  $\Delta G$  value for our four zinc finger PRDM9<sub>A</sub> ( $\Delta G_{MM/PBSA} = -74.66$ ) and the three zinc finger 1AAY ( $\Delta G_{MM/PBSA} = -18.38$  Kcal/mol), it showed a difference in  $\Delta G$  with  $\Delta\Delta G$  equal to  $-56.28$  kcal/mol in favor of PRDM9<sub>A</sub>. This result indicates that the affinity of the ZnF to its target DNA increases as the number of zinc fingers in the protein increase. This disagree with the report by Pabo which state that: as the number of fingers connected by canonical linkers increases, the binding constant started to plateau for four fingers and more. No additional stability was observed with four-fingers and more<sup>12</sup>.

Both 1AAY and 5EGB have the same zinc finger pattern of  $X_2CX_{2-4}CX_{12}HX_{2-8}H$ . The established binding trend that three  $\alpha$  helices of 1AAY bind to the major groove of the target DNA through the surface amino acids side chains at position -1, 2, 3 and 6<sup>12</sup>. Residues at -1, 3, 6 positions (R, E, and R, respectively) bind to three bases of the primary strand and residue at position 2 (D) binds to the fourth base in the complementary strand. But in PRDM9<sub>A</sub> binding to DNA most of the binding take place in the major groove of the complementary strand (3' to 5') at -2 (S or R), -1 (N or D), 3 (H or N), 6 (R or S) positions that lie on the inside face of the helix or in the preceding loop<sup>26</sup>.

#### 3.1.1.4 Alleles of PRDM9 Protein Binding to DNA

The interaction of amino acids and DNA nucleotides are structurally diverse. Some researchers established databases such as Protein-Side Chain Interaction database<sup>69,70</sup>. This data base categorizes the amino acid-nucleotide interactions. For example, Arg and His are more specific for guanosine (G), Lys and Ser are more specific for thymidine (T), Glu is more specific for cytidine (C) and Asp is more specific for adenosine and thymidine (A,T)<sup>68-70</sup>.

The change in the affinity of binding among different alleles are shown in Table 3.5. *Allele LI3* showed a higher affinity than *allele A* with  $\Delta\Delta G_{MMPBSA} = -16.53$  kcal/mole due to the change of serine (S at position 140) to positively



charged arginine (R) which is Known to be more specific to G than Ser<sup>29,68-70</sup> (Figure 3.6).

*Allele L9/L24* showed a lower affinity than *allele A* with  $\Delta\Delta G_{MM/PBSA} = +0.89$  kcal/mole due to the change of positively charged lysine (K) to a negatively charged glutamate (E). As a result, a repulsion occurs instead of the electrostatic interaction which reduces the affinity of the protein to the DNA. *Allele L9* from *allele L24* have different but synonymous base substitutions for the same zinc finger array<sup>26</sup>.

*Allele L20* showed a higher affinity than *allele A* with  $\Delta\Delta G_{GMM/PBSA} = -4.67$  kcal/mole and  $\Delta\Delta G_{MM/GBSA} = 4.96$  kcal/mole, but experimentally it showed a lower affinity with  $\Delta\Delta G = 1.0$  kcal/mole. MM/GBSA agree with the experimental change while MM/PBSA, in this case, disagrees with the experimentally found values. *Allele L20* has a positively charged Histidine in position 116 (H116) which is specific to G and C. While *allele A* has a negatively charged Asparagine in position 116 (specific to A). It is worth noting that this position in the wild type specific to A and this replacement reduced the affinity. Upon substituting T:A with C:G in the DNA sequence (Figure 3.7). the affinity for allele L20 decreased with a  $\Delta\Delta G_{MM/GBSA}$  and  $\Delta\Delta G_{MM/PBSA}$  equal +12.47 and +25.1 kcal/mol, respectively. While in experimental study, this change increases the affinity for the *allele L20*

with a  $\Delta\Delta G = -0.5 \text{ kcal/mol}^{26}$ . These calculated values are in agreement with the experimental changes within error.

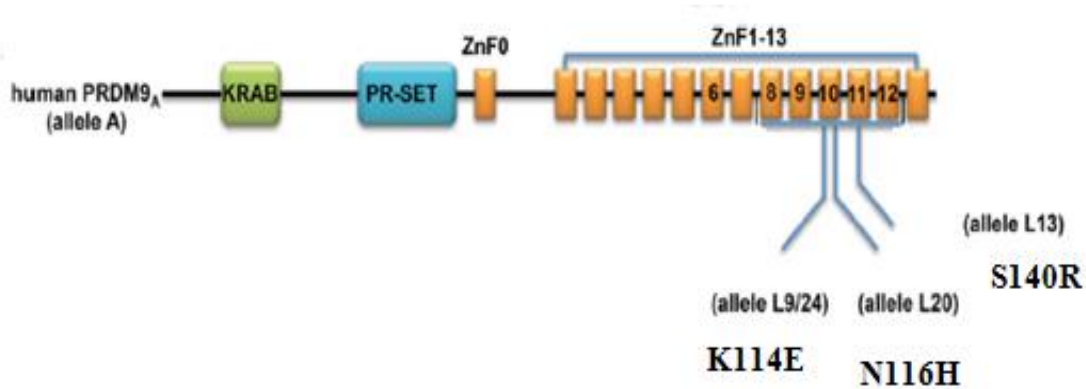


Figure 3.6: The non-A alleles of hPRDM9 where a single amino acid substitution at ZnF3 and ZnF4 resulted in alleles L9/L24, L20 or L13, respectively<sup>26</sup>.

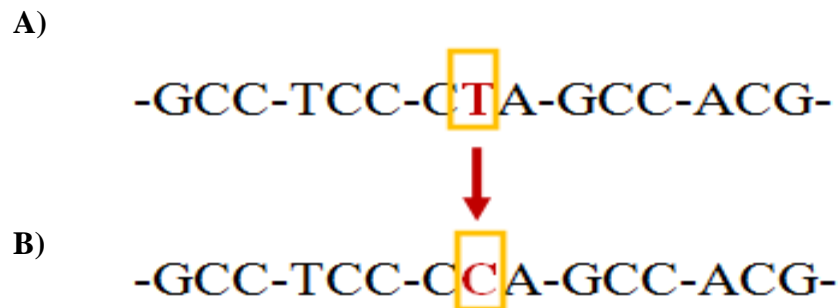


Figure 3.7 : A) DNA strand (5`-3`) of the PARMD9<sub>L20</sub>-DNA complex. B) Mutant DNA strand (5`-3`) of the PARMD9<sub>L20</sub>-DNA complex.

Table 3.5. Comparison of theoretically calculated  $k_d$  values from both MM/GBSA and MM/PBSA with the experimentally derived values using Gel-shift assays.

Complexes	Calculated				Experimental	
	$\Delta G_{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G_{\text{binding}} - T\Delta S$	$\Delta G_{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G_{\text{binding}} - T\Delta S$	$k_d$ $=10^{(\Delta G_{\text{MM/GBSA}}/1.4)}$ nM	$k_d$ $=10^{(\Delta G_{\text{MM/PBSA}}/1.4)}$ nM	$k_d$ nM	$\Delta G$ $=1.4 \log K_d(\text{M})$ kcal/mol
<b>1AAY</b> (3F)	+4.270	-18.38	$11.20 \cdot 10^2$	$7.4 \cdot 10^{-5}$	13	-11.04
<b>5EGB</b> (4F)	-29.48	-74.66	$8.770 \cdot 10^{-13}$	$4.69 \cdot 10^{-45}$	36	-10.40
<b>L9/24</b> (4F)	-31.92	-73.77	$1.585 \cdot 10^{-14}$	$2.03 \cdot 10^{-44}$	494	-8.829
<b>L13</b> (4F)	-27.83	-91.19	$1.321 \cdot 10^{-11}$	$1.96 \cdot 10^{-57}$	25	-10.60
<b>L20<sub>TA</sub></b> (4F)	-22.44	-79.33	$9.354 \cdot 10^{-8}$	$2.16 \cdot 10^{-48}$	211	-9.346
<b>L20<sub>CG</sub></b> (4F)	-9.970	-54.20	75.60	$1.93 \cdot 10^{-30}$	91	-9.857

Table 3.4. Calculated energies for PRDM9<sub>A</sub>-DNA complex and its different alleles.

Protein-DNA Complex	Mutation	Specificity	$K_d \text{ exp}$ "nM"	$\Delta G_{\text{exp}}$ "Kcal/mol"	$\Delta G_{\text{calc}} \text{ "GB"}$ "Kcal/mol"	$\Delta G_{\text{calc}} \text{ "PB"}$ "Kcal/mol"
<b>5egb</b>	F10 (K114,N116), F11 (S140)	K spec. to G N spec. to A,C S spec. to T	36	0-10.4	-29.48	-74.66
<b>L9/24</b>	F10 (K114E)	E spec. to T	494	29-8.8	-31.92	-73.77
<b>L13</b>	F11 (S140R)	R spec. to G	25	-10.60	-27.83	-91.19
<b>L20<sub>TA</sub></b>	F10 (N116H)	H spec. to G,C	211	-9.346	-22.44	-79.33
<b>L20<sub>CG</sub></b>	T:A to C:G	H spec. to G,C	91	-9.857	-9.970	-54.20

### 3.1.1.5 PRDM9<sub>A</sub>- DNA Interaction: Hydrogen-Bonding

The major contacts of PRDM9<sub>A</sub> (four zinc finger protein) were reported to take place in the DNA major groove<sup>26</sup>. Most of the Hydrogen bond contacts take place with purine bases (G or A) in the complementary strand (3`-5`). Residues at position -2(S or R) , -1(N or D) , 3 (H or N), 6 (R or S) vary in their degree of contact to the DNA according to the finger involved in binding i.e whether in F1 or F2 or F3 or F4<sup>26</sup>. Zinc fingers 1, 2 and 4 contact the guanines (G) through H bond with conserved histidine (H) residues at position 3 or arginine (R) residues at position 6. NH groups of Arg63, Arg91 and Arg147 donate H bond to O6 and N7of guanine (G) (Figure 3.8). Terminal NH groups of His60, His88 and His144 donate one Hydrogen bond to O6 or N7 of guanine (Figure 3.9).

In this work the four  $\alpha$ -helices (fingers) of PRDM9<sub>A</sub> zinc finger protein formed hydrogen bond contacts with three to four adjacent bases on the DNA. This finding disagrees with the report that only three zinc fingers bind the DNA irrespective of the number of fingers of the protein<sup>8,12</sup>. To the contrary, it is evident that upon increasing the number of zinc fingers, the affinity increases to the target DNA, and more than three fingers are bound in this case.

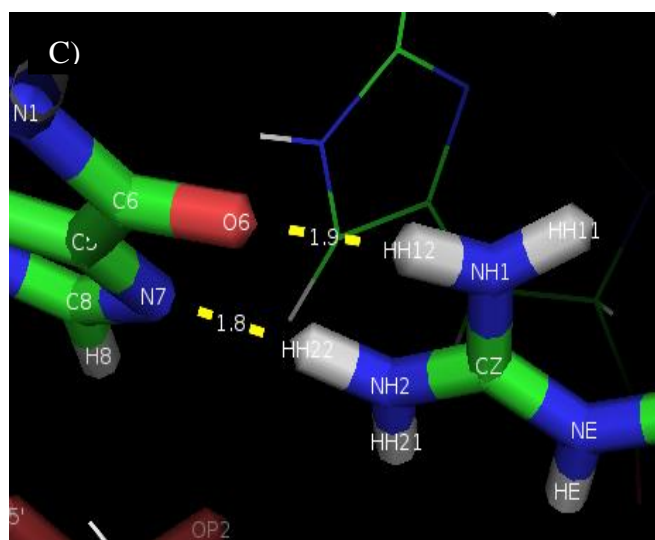
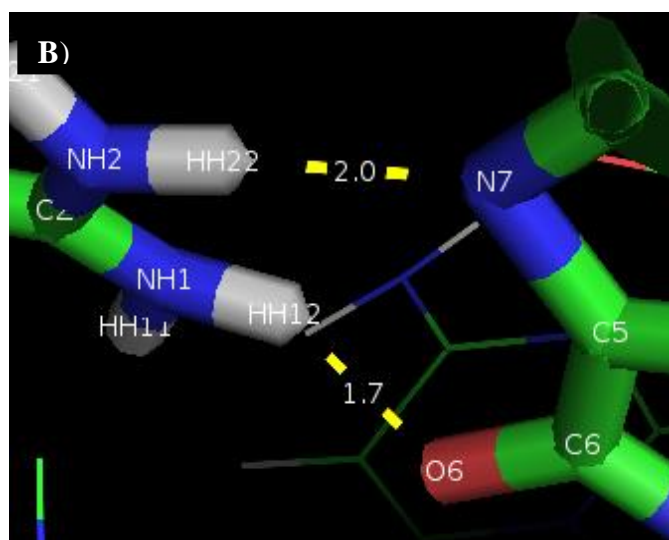
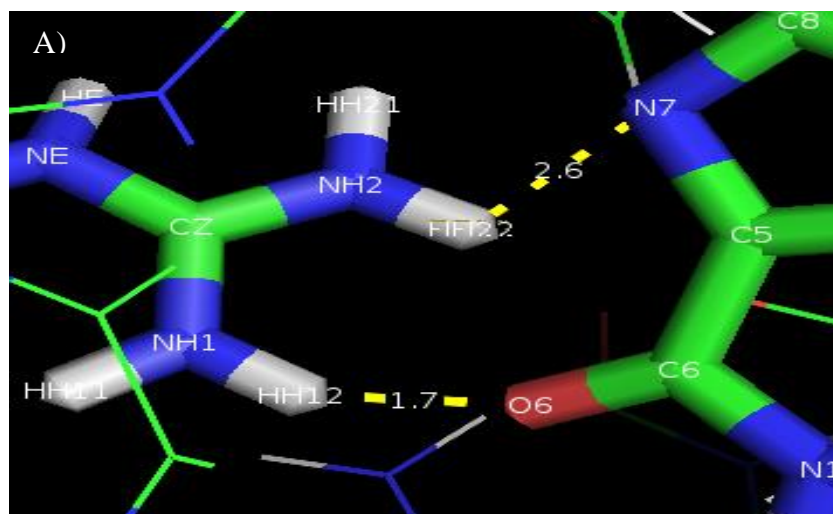


Figure 3.8: 5EGB-DNA contacts: Guanine H-bond contacts with: A) Arg63, B) Arg91 and C) Arg147 residues at position 6 with inter-atomic distances in F1, F2, F4.

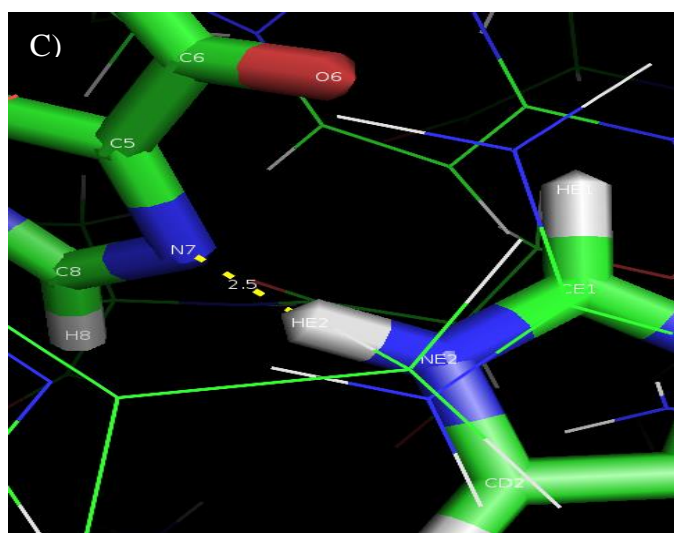
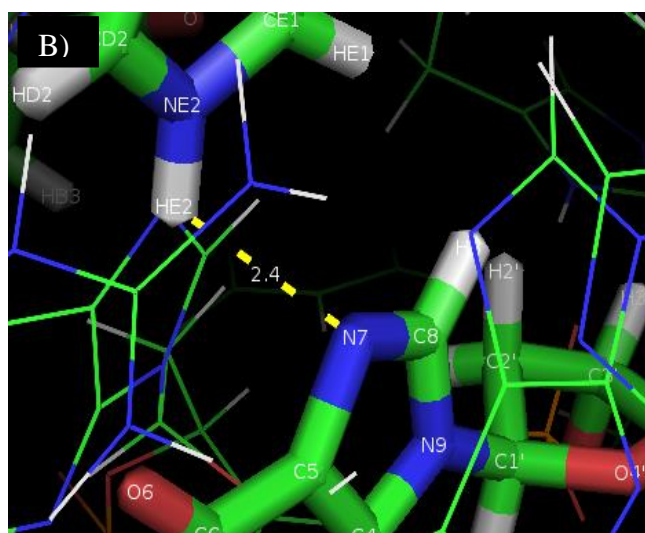
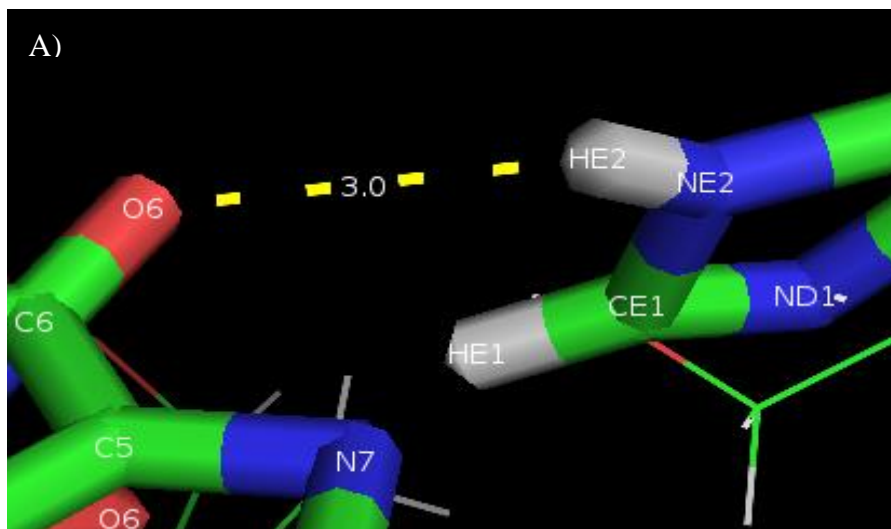


Figure 3.9: 5EGB-DNA contacts: Guanines H-bond contacts with: A) His60, B) His88 and C) His144 residues at position 3 with interatomic distances in ZNF1, 2, 4.

However, *Finger 3* contains Asn116 at position 3 and Ser119 at position 6. The side chain of Asn116 donates one H bond to N7 of adenine (A) and accepts H bond from N6 of adenine (A). The partner thymine (T) interacts nonspecifically with a water molecules through O4` (Figure 3.10).

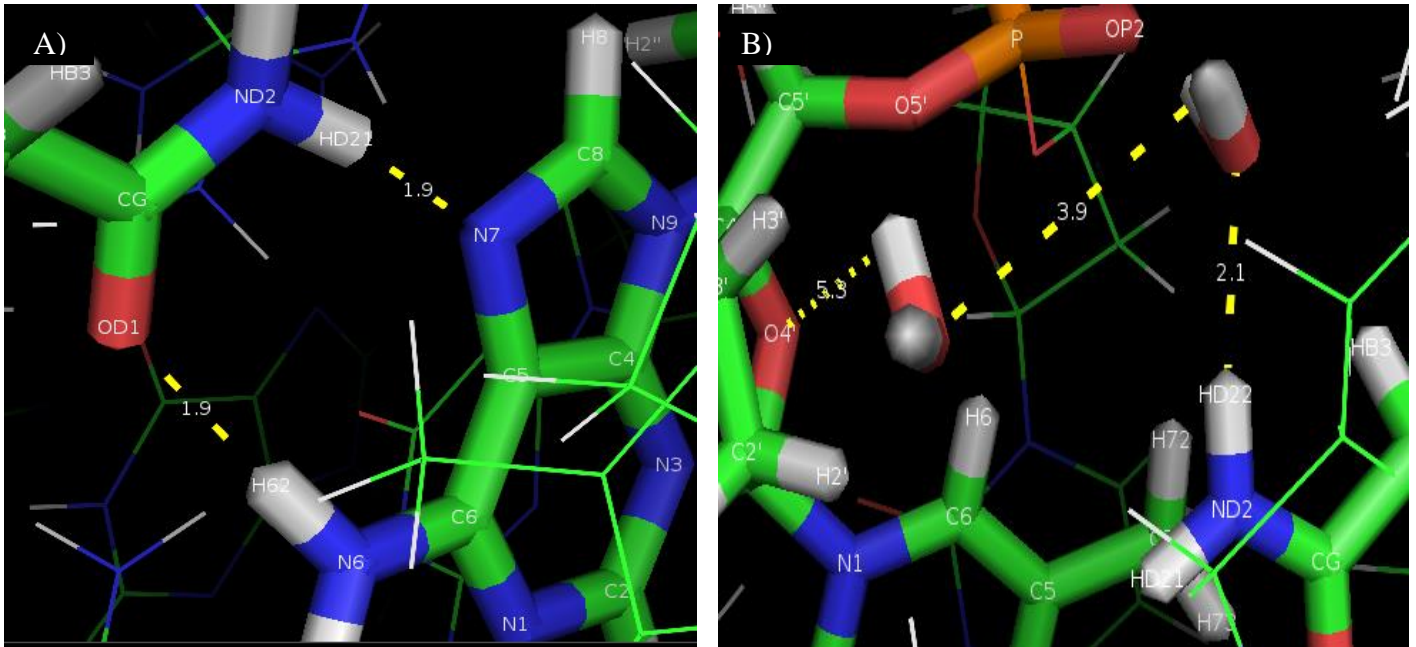
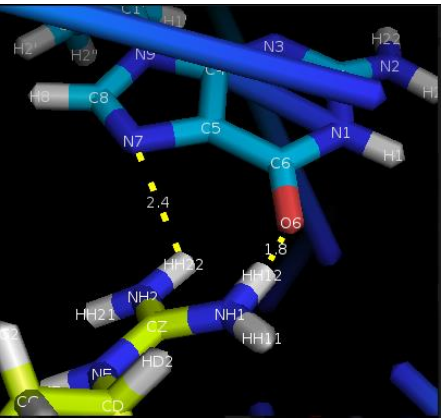
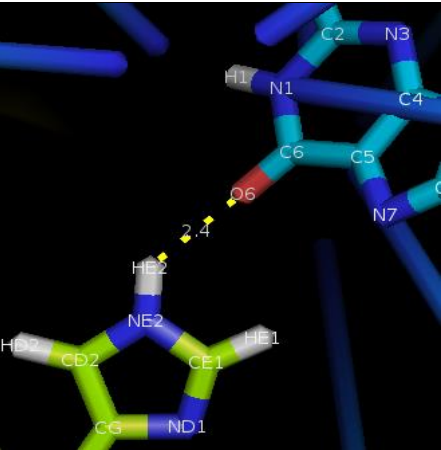
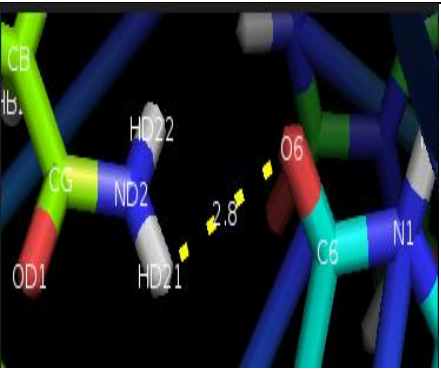


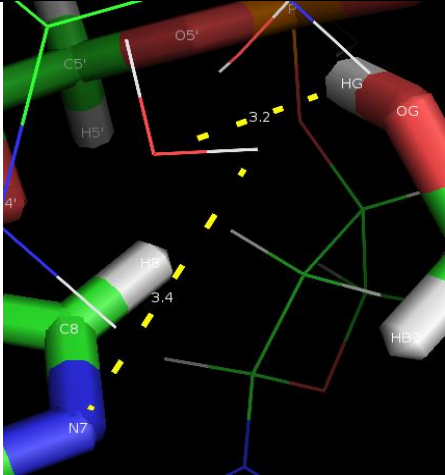
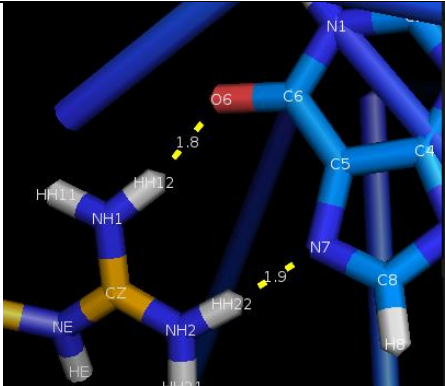
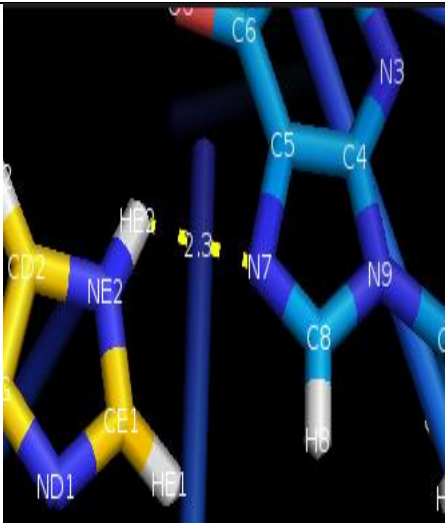

Figure 3.10: 5EGB-DNA contacts: Adenine and Thymine H-bond contacts with Asn116 residues at position 3 in F3: A) Adenine contact and B) Thymine water mediated nonspecific contact.

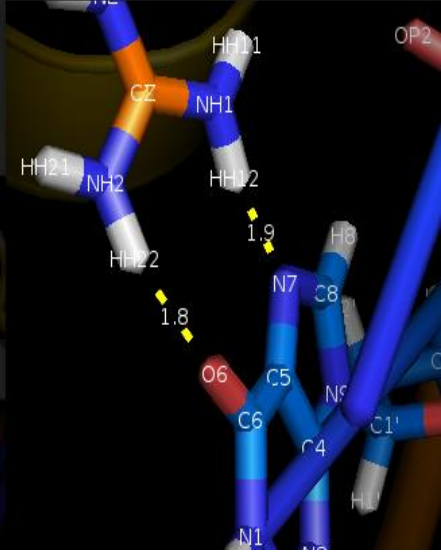
Ser119 residues at position 6 in F3 interact with thymine (T 3`-5`) nonspecifically with O4 from water molecules. The water mediated H-bond is a result of the short side chain of Ser. Adenine (A 5`-3`) opposite to T contacts residue Ser188 at position 2 in F4. The water molecules mediated contacts between residues side chains and base pairs are important for specific recognition and enhancing the affinity and specificity of the of the protein to the DNA sequence<sup>26</sup>.

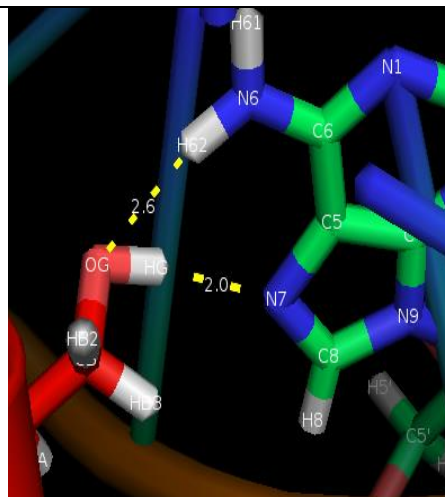
Table 3.7: The direct hydrogen bond contacts between 5EGB-DNA complex.

Zinc Finger	Acceptor	Donor	Distance Å	PYMOL Screen Shot
<b>F1</b>	G <sub>17</sub> -O <sub>6</sub>	Arg <sub>63</sub> -HH <sub>12</sub>	1.8	
	G <sub>17</sub> -N <sub>7</sub>	Arg <sub>63</sub> -HH <sub>22</sub>	2.4	
	G <sub>18</sub> -O <sub>6</sub>	His <sub>60</sub> -HE <sub>2</sub>	2.4	
G <sub>25</sub> -O <sub>6</sub> (5'-3')	Asn <sub>57</sub> -HD <sub>21</sub>	2.8		



	G <sub>25</sub> -N <sub>7</sub> (5'-3')	Ser <sub>59</sub> -HG	Water mediated	
<b>F2</b>	G <sub>14</sub> -O <sub>6</sub>	Arg <sub>91</sub> -HH <sub>12</sub>	1.8	
	G <sub>14</sub> -N <sub>7</sub>	Arg <sub>91</sub> -HH <sub>22</sub>	1.9	
	G <sub>15</sub> -N <sub>7</sub>	His <sub>88</sub> -HE <sub>2</sub>	2.3	
	A <sub>12</sub> -N <sub>7</sub>	Asn <sub>116</sub> -HD <sub>21</sub>	1.9	

<b>F3</b>	Asn <sub>116</sub> -OD <sub>1</sub>	A <sub>12</sub> -H <sub>62</sub>	2.0	
	G <sub>13</sub> -O <sub>6</sub>	Arg <sub>112</sub> -HH <sub>22</sub>	1.8	
	G <sub>13</sub> -N <sub>7</sub>	Arg <sub>112</sub> -HH <sub>12</sub>	1.9	
	A <sub>33</sub> -N <sub>7</sub> (5'-3')	Ser <sub>143</sub> -HG	2.0	
	Ser <sub>143</sub> -OG	A <sub>33</sub> -H <sub>62</sub> (5'-3')	2.6	



	T <sub>11</sub> -O <sub>4</sub>	Asn <sub>116</sub> -HD <sub>22</sub>	Water mediated	
	T <sub>32</sub> -O <sub>4</sub>	Ser <sub>119</sub> -HG	Water mediated	
<b>ZnF4</b>	G <sub>8</sub> -O <sub>6</sub>	Arg <sub>147</sub> -HH <sub>12</sub>	1.8	

	G <sub>9</sub> -N <sub>7</sub>	His <sub>144</sub> -HE <sub>2</sub>	2.4	
	Asn <sub>141</sub> -OD <sub>1</sub>	C <sub>10</sub> -H <sub>42</sub>	2.7	

### 3.1.2 Binding of PPR1 and Its Mutants to DNA

From the MD simulations, the RMSD for the complex backbone (CA, C, N) has been plotted (Figure 3.11). It shows that RMSD values fluctuated steadily around 2.7 Å for PPR1-DNA complex indicating reasonable stability upon simulation<sup>65</sup>.

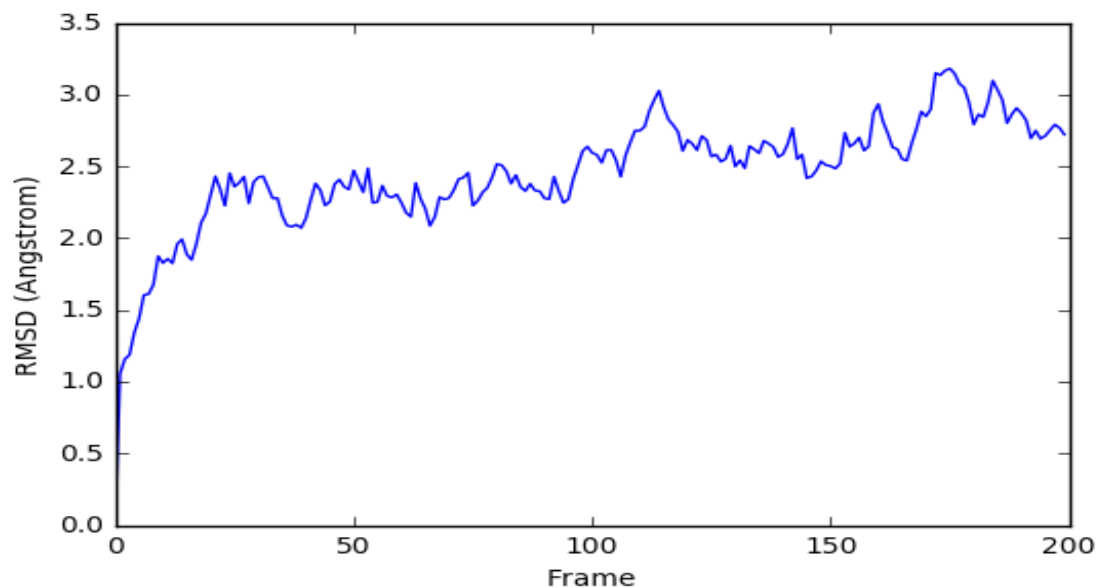
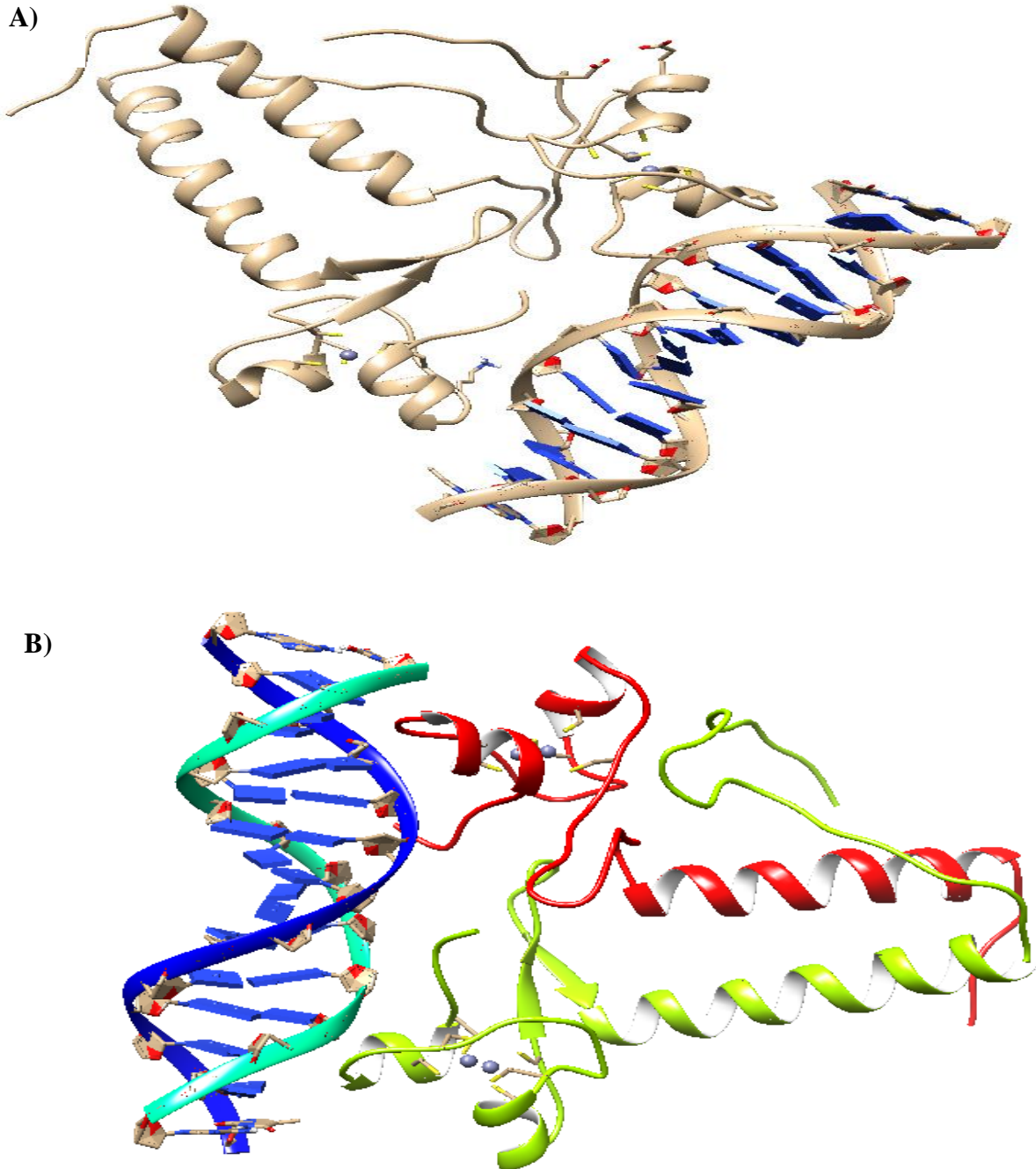


Figure 3.11: Backbone (CA, C, N) RMSD vs. Frame (reference residue number) after the production phase of the MD refinement for 1PYI.

#### 3.1.2.1 Overall Structure of PPR1-DNA Complex

PPR1 is a dimer zinc finger protein which binds DNA. The average structure of the PPR1 complex with DNA from molecular modeling simulations is shown in Figure 3.12<sup>38</sup>. PPR1 contains six cysteine residues, two zinc ions. This Zn<sub>2</sub>Cys<sub>6</sub> domain binds to 14 base pairs recognition sites as a nonsymmetrical homo-dimer. The two Zn<sub>2</sub>Cys<sub>6</sub> domain lie above the major groove and therefore interact with the

CGG sites near each end of the DNA. The  $Zn_2Cys_6$  domain contains two short helices. Each helix is followed by a turn, and an extended strand (Figure 3.13).



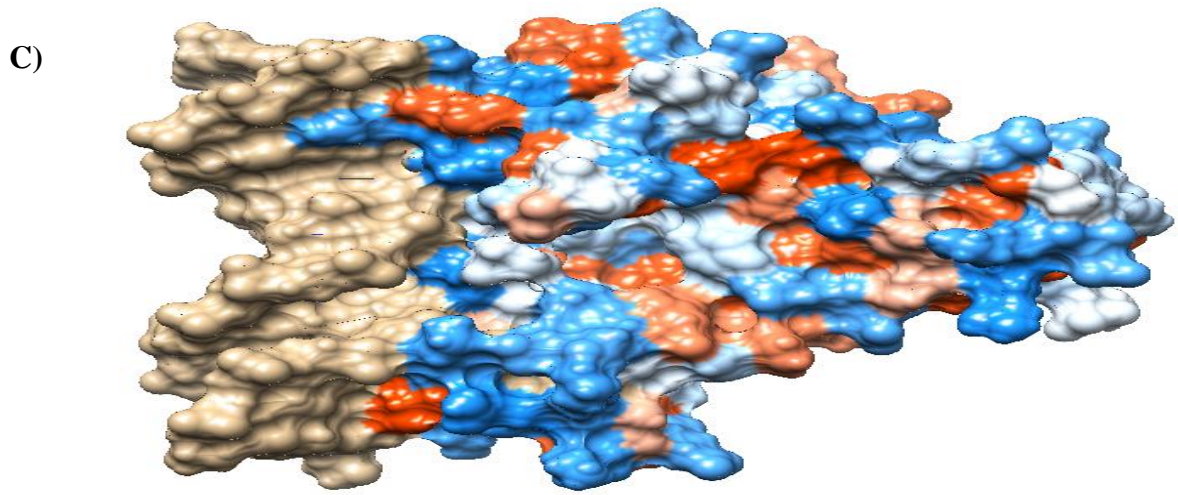


Figure 3.12: Overall structure of PPR1-DNA complex after molecular modeling simulation, showing the two binuclear cluster ( $Zn_2Cys_6$ ). A) PPR1 complex with the DNA strands, two binuclear cluster are represented as a light brown and the coordinated zinc ions are represented as blue spheres. B) The PPR1-DNA complex are represented as a rounded ribbon, first binuclear cluster are red, the second one are green the coordinated zinc ions are represented as a blue spheres. C) The PPR1 complex are represented as a hydrophobicity surface, DNA strands are a light brown.

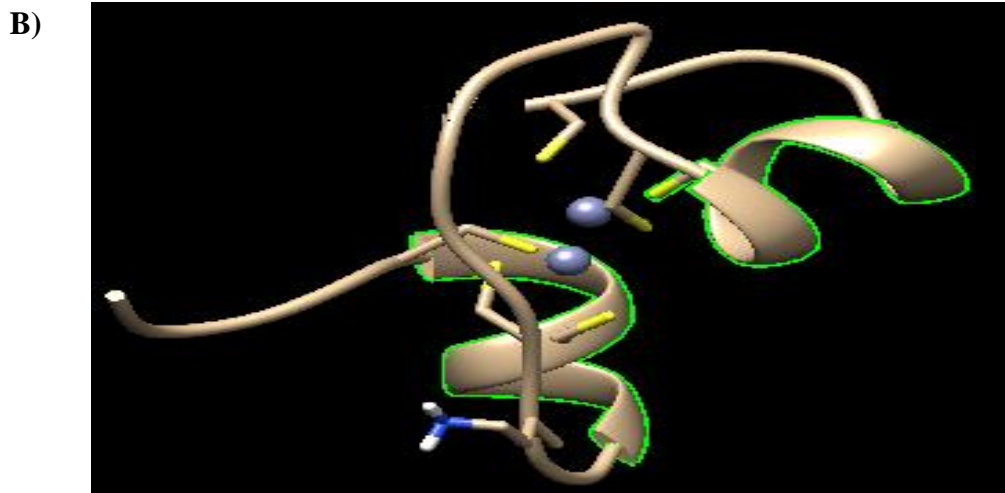
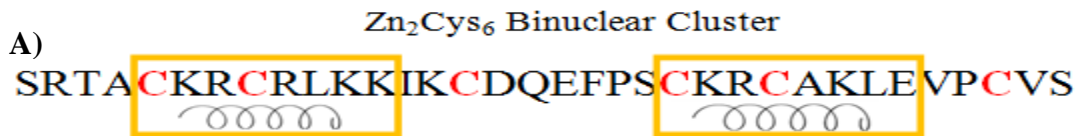


Figure 3.13: A) Amino acid sequence of the Zn domain of the PPR1. B) Structure of the Zn domain by Chimera showing two short helices and the coordinated zinc ions as blue spheres.



### 3.1.2.2 Binding of PPR1 Compared to the three Zinc Finger Protein 1AAY to DNA

1AAY-DNA binding domain has three zinc fingers that bind the major groove in a tandem arrangement (Figure 3.14). Each finger binds 3-4 base pairs and most of the base contacts are to the primary strand of the DNA. Residues at position -1, 2, 3, 6 in each  $\alpha$  helix of 1AAY make specific base contacts<sup>12</sup>.

PPR1 contains a domain of six cysteine residues and two zinc ions ( $Zn_2Cys_6$  binuclear cluster). The homo-dimer PPR1 is bound asymmetrically to the major groove of the DNA target that contains two inverted CGG half sites separated by a 6 base pair through the binuclear cluster region (Figure 3.12). The base contacts were to both strands of the DNA (primary and complementary strands). Lys residue at position 40 and 41 in both nuclear clusters make a series of interactions with the CGG bases near each end of the DNA<sup>38</sup> (Figure 3.15).

$\Delta G_{MM/GBSA}$  and  $\Delta G_{MM/PBSA}$  for PPR1 protein binding was -7.35, -18.52 kcal/mol, respectively. For the binding of 1AAY to DNA the  $\Delta G_{MM/PBSA}$  was -18.38 kcal/mol. It is observable that the homo-dimer PPR1 has a slightly stronger binding than three zinc finger 1AAY within margin of error in MM/PBSA, and a weaker interaction in the MM/GBSA results (Table 3.8, Table 3.9).



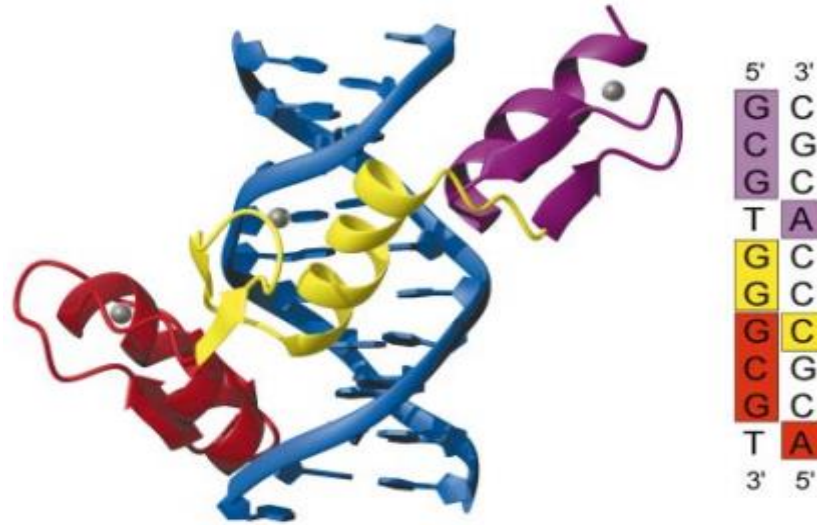


Figure 3.14: The Zif268-DNA complex showing the three zinc fingers of Zif268 bound in the major groove of the DNA. Fingers are spaced 3-base pairs intervals. The DNA is blue; fingers 1, 2, and 3 of Zif268 are red, yellow, and purple, respectively; and the coordinated zinc ions are represented as silver spheres. The DNA sequence of Zif268 site on the right is color-coded to indicated base contacts made by each finger<sup>12</sup>.

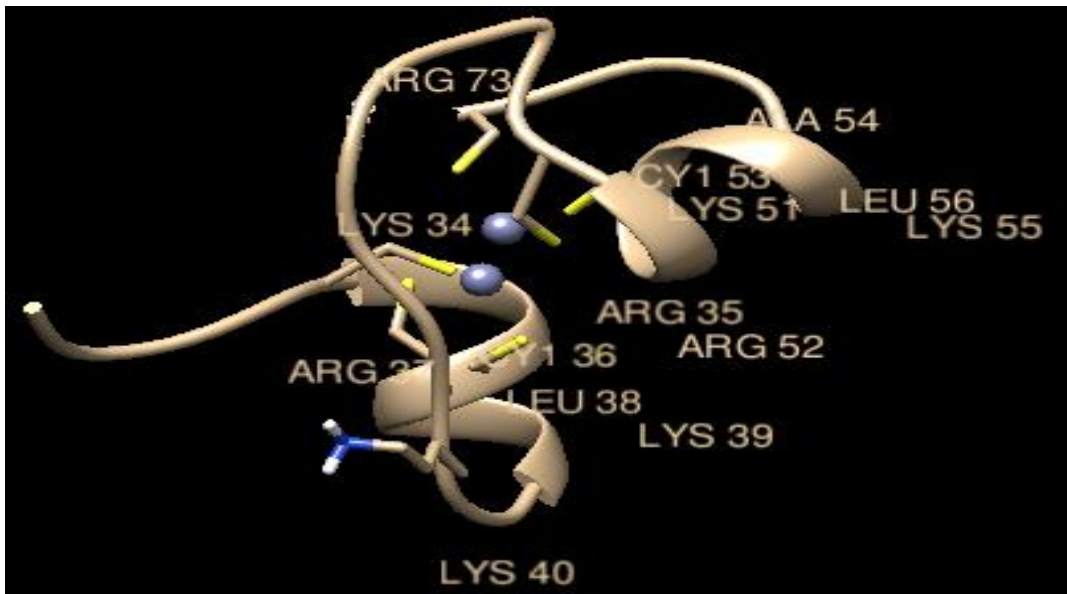


Figure 3.15: The Binuclear cluster region of PPR1 after molecular dynamic simulations, showing the coordinated zinc ions as blue spheres and the Lys residue at position 40 and 41 in both nuclear clusters that make a series of interactions with the CGG bases near each end of the DNA as light brown.

Table 3.8. Calculated energies (kcal/mol) for 1AAY and 1ZME DNA complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM/GBSA and MM/PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting  $T\Delta S$  from *nmode* calculations from the binding free energy.

Complexes	$\Delta G^{\text{binding}}$ ( $\Delta H$ ) (MM/GBSA) kcal/mol	$\Delta G^{\text{binding}}$ ( $\Delta H$ ) (MM/PBSA) kcal/mol	$T\Delta S$ kcal/mol	$\Delta G^{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$	$\Delta G^{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$
<b>1AAY</b> <b>(3F)</b>	-84.09	-106.74	-88.36	+4.270	-18.38
<b>1PYI</b> (Binuclear cluster)	-74.75	-85.92	-67.40	-7.350	-18.52
<b>1PYI mutant</b> (Binuclear cluster)	-35.01	-39.99	-56.81	+21.80	+16.82

Table 3.9. Comparison of theoretically calculated  $k_d$  values from both MM/GBSA and MM/PBSA with the experimentally derived values using Gel-shift assays.

Complexes	Calculated				Experimental	
	$\Delta G^{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$	$\Delta G^{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G^{\text{binding}}-T\Delta S$	$k_d$ $=10^{(\Delta G^{\text{MM/GBSA}}/1.4)}$ nM	$k_d$ $=10^{(\Delta G^{\text{MM/PBSA}}/1.4)}$ nM	$k_d$ nM	$\Delta G$ $=1.4\log K_d(M)$ kcal/mol
<b>1AAY</b> <b>(3F)</b>	+4.270	-18.38	$11.2 \cdot 10^{12}$	$7.4 \cdot 10^{-5}$	13	-11.04
<b>1PYI</b> (Binuclear cluster)	-7.350	-18.52	$5.6 \cdot 10^3$	$5.9 \cdot 10^{-5}$	7	-11.41

### 3.1.2.3 Binding Affinities of PPR1-DNA Complex and its Mutant

The absolute energy of PPR1 protein calculated using MM/GBSA method was less in negative than the experimental value ( $\Delta G = -11.41$  kcal/mol) with a  $\Delta\Delta G = -4.06$  kcal/mol. On the other hand, the absolute energy calculated using MM/PBSA method was higher in the negative than the experimental value with a  $\Delta\Delta G = +7.11$  kcal/mol (Table 3.9).

In Vitro binding studies of PPR1 protein show that PPR1 proteins are insensitive to mutation of the base pair separating DNA half sites. However, they are extremely sensitive to the number of base pair separating the CGG half site<sup>44</sup>. To test this finding a mutation in the CGG half site themselves was applied as shown in Figure 3.16.



Figure 3.16: PPR1-Binding sites containing two highly CGG elements near each end of the DNA. A) without mutation. B) with mutation at the CGG half sites.

A mutant of PPR1 (CGG at half sites changed to TAA) gave a highly positive absolute energy as calculated by both the MM/GBSA and MM/PBSA methods with a  $\Delta\Delta G_{MM/GBSA}$  and  $\Delta\Delta G_{MM/PBSA}$  absolute equal 29.15, 35.34 kcal/mole, respectively relative to the wild type PPR1 protein (Table 3.8). This result indicate that the dissociation is preferred over the association<sup>8</sup>, higher affinity to the wild type DNA site. The mutation in the CGG binding site destroyed most of the binding of the PPR1 protein to the DNA, so we can safely conclude that PPR1 protein is more specific to the CGG site.

#### **3.1.2.4 PPR1- DNA Interaction: Hydrogen-Bonding**

Both amino-terminal Zn domains ( -C<sub>34</sub>K<sub>35</sub>R<sub>36</sub>C<sub>37</sub>R<sub>38</sub>L<sub>39</sub>K<sub>40</sub>K<sub>41</sub>- ) bind to a 14 base pair recognition DNA site with specific CGG triplets symmetrically placed near each end as a nonsymmetrical homo-dimer in the major groove<sup>38</sup>. None of the residues in the linker region and the coiled-coil make contacts with the base nucleotide. Instead they make contacts with the sugar phosphate backbone only.

Most of the base pair hydrogen bonding that was found in our simulation was similar to the hydrogen bonding found by experiment (Figure 3.17). Most of the contacts were through the Lys at position 40 in both binuclear cluster and Lys at position 41 (Figure 3.17).

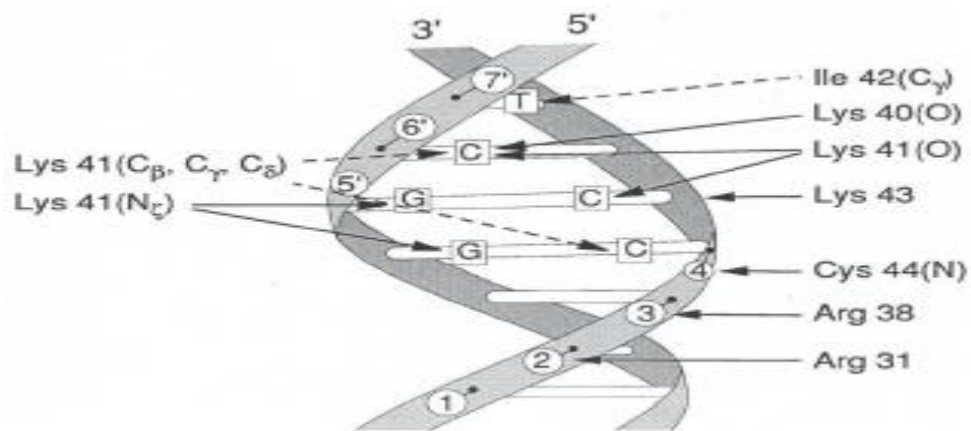


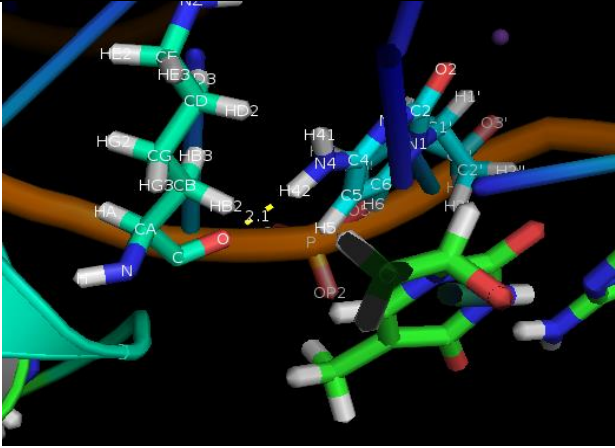
Figure 3.17: Schematic diagram showing all phosphate and base contacts between 1PPR1 protein and DNA strand<sup>38</sup>.

Table 3.10. The direct contacts between both binuclear clusters  $C_{34}K_{35}R_{36}C_{37}R_{38}L_{39}K_{40}K_{41}$  and the 5'-3' sequence and its 3'-5' complementary sequence. The numbering of protein sequence is less by one than the X-ray structural ( $K_{40} = K_{39}$  in our case).

Acceptor	Donor	Distance Å	PYMOL Screen Shot
Lys <sub>39</sub> -O	DC <sub>2</sub> -H <sub>42</sub>  5'-3'	2.3	

<p>DG<sub>3</sub>-N<sub>7</sub></p> <p>5`-3`</p>	<p>Lys<sub>40</sub>-HZ<sub>2</sub></p>	<p>2.1</p>	
<p>DG<sub>4</sub>-O<sub>6</sub></p> <p>5`-3`</p>	<p>Lys<sub>40</sub>-HZ<sub>3</sub></p>	<p>2.0</p>	
<p>Lys<sub>40</sub>-O</p>	<p>DC<sub>26</sub>-H<sub>42</sub></p> <p>3`-5`</p>	<p>2.1</p>	

Lys <sub>127</sub> -O	DC <sub>16</sub> -H <sub>42</sub>  3`-5`	1.9	
DG <sub>17</sub> -N <sub>7</sub>  3`-5`	Lys <sub>128</sub> -HZ <sub>2</sub>	2.1	
DG <sub>18</sub> -O <sub>6</sub>  3`-5`	Lys <sub>128</sub> -HZ <sub>3</sub>	1.8	

Lys <sub>128</sub> -O	DC <sub>12</sub> -H <sub>42</sub>  5`-3`	2.4	
-----------------------	--	-----	--

### 3.1.3 Binding of PUT3 and Its Mutants to DNA

From the MD simulations, the RMSD for the complex backbone (CA, C, N) has been plotted (Figure 3.18). It shows that RMSD values fluctuated steadily around 2.5 Å for PUT3-DNA complex indicating reasonable stability upon simulation<sup>65</sup>.

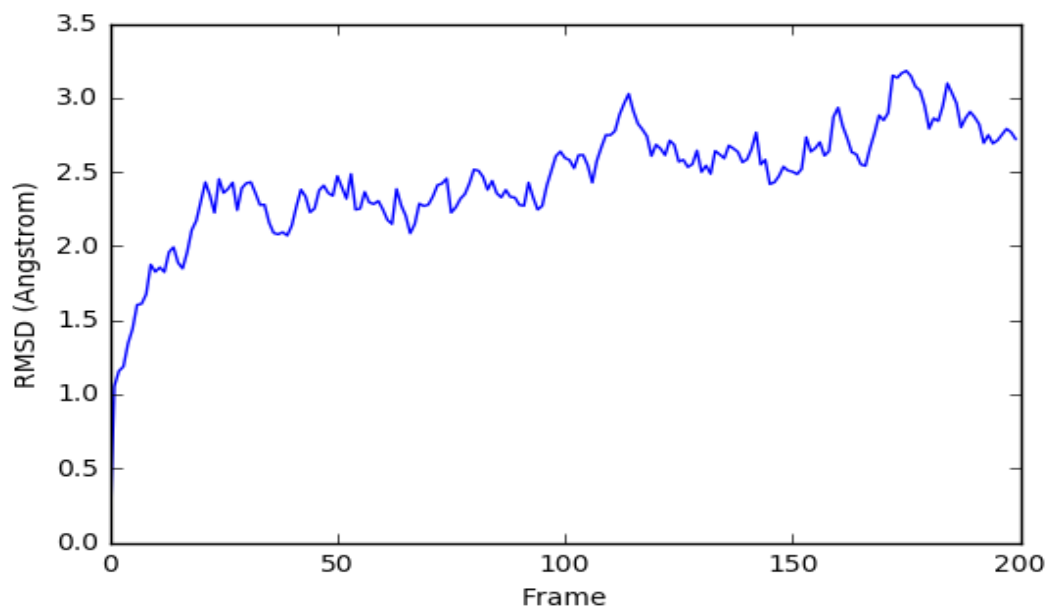
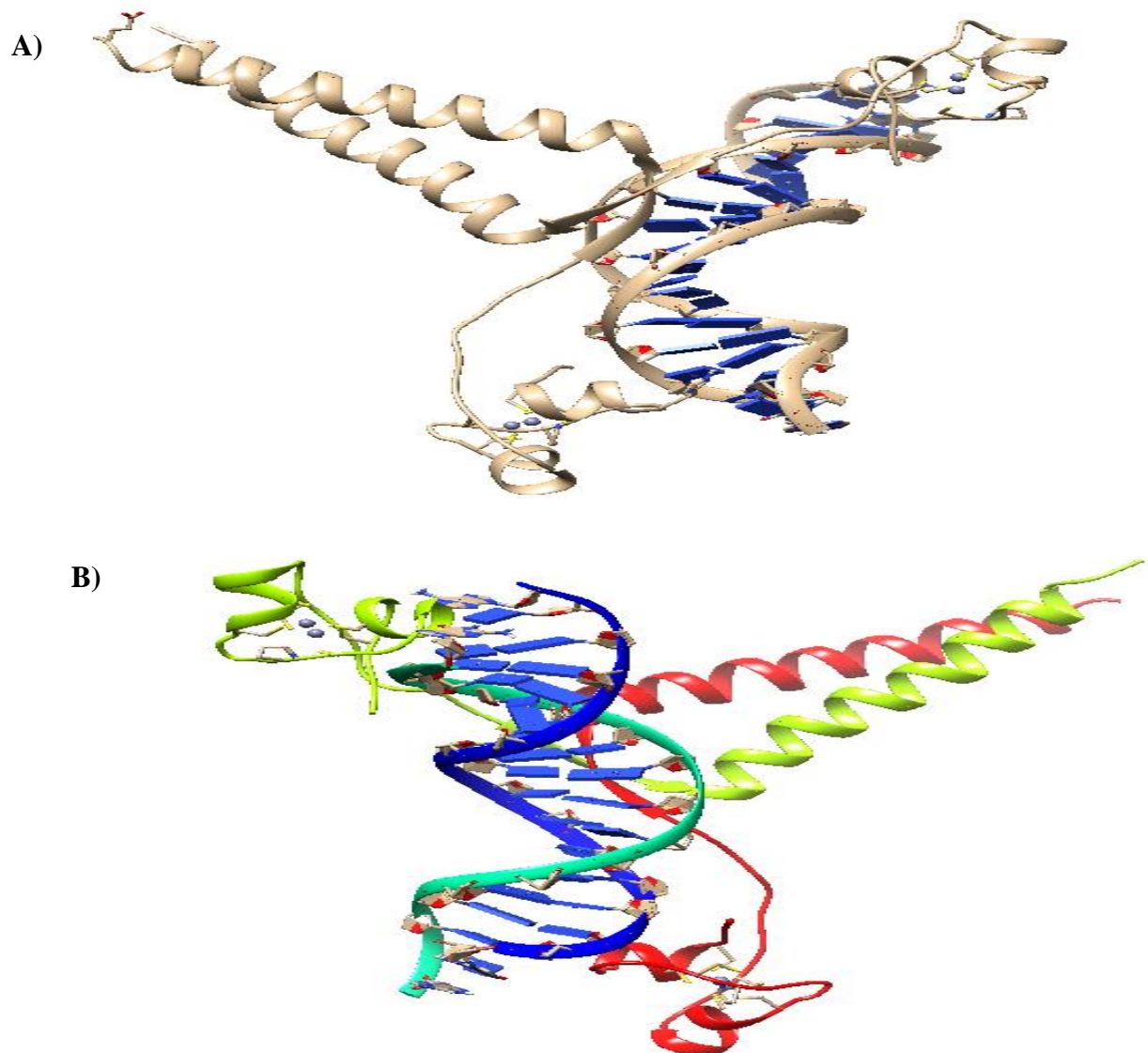


Figure 3.18: Backbone (CA, C, N) RMSD vs. Frame (reference residue number) after the production phase of the MD refinement for PUT3.



### 3.1.3.1 Overall Structure of PUT3-DNA Complex

The average structure of the complex PUT3-DNA from molecular modeling simulations is shown in Figure 3.19. The homo-dimer PUT3 bound asymmetrically to the DNA. The protein dimer is turned around the DNA by one and a one-half turns similar to what was found by X-ray<sup>44</sup>. The two  $Zn_2Cys_6$  domains contact the CGG sites and are symmetrically disposed and they lie in the major groove of the DNA at each end.



C)

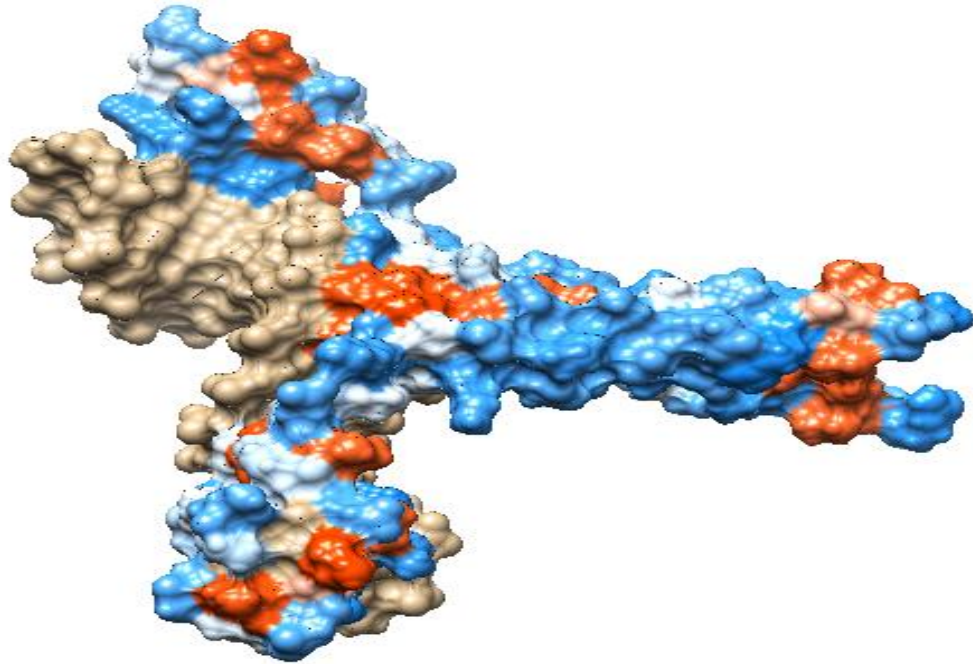


Figure 3.19: Overall structure of PUT3-DNA complex after molecular modeling simulation, showing the two binuclear cluster ( $Zn_2Cys_6$ ). A) PUT3 complex with the DNA strands, two binuclear cluster are represented as a light brown and the coordinated zinc ions are represented as blue spheres. B) The PUT3-DNA complex are represented as a rounded ribbon, first binuclear cluster are red, the second one are green the coordinated zinc ions are represented as a blue spheres. C) The PUT3 complex are represented as a hydrophobicity surface, DNA strands are a light brown.

### 3.1.3.2 Binding of PUT3 Compared to the three Zinc Finger Protein 1AAY to DNA

1AAY-DNA binding domain has three zinc fingers that bind in the major groove in a tandem arrangement (Figure 3.20). Each finger binds 3-4 base pairs and most of the base contacts are to the primary strand of the DNA. Residues at position -1, 2, 3, 6 in each  $\alpha$  helix of 1AAY make specific base contacts<sup>12</sup>.

PUT3 contains a domain of six cysteine residues and two zinc ions ( $Zn_2Cys_6$  binuclear cluster). The homo-dimer PUT3 is bound asymmetrically to the major

groove of the DNA target that contains two inverted CGG half sites separated by a 10 base pair through the binuclear cluster region (Figure 3.19). The base contacts were to both strands of the DNA (primary and the complementary strands). Residues at position 42, 44 and 45 in both nuclear clusters responsible for the contacts to the CGG end of the DNA<sup>44</sup> (Figure 3.21).

$\Delta G_{MM/GBSA}$  for PUT3 binding was -15.79 kcal/mol while the  $\Delta G_{MM/PBSA}$  for binding of 1AAY to DNA was -18.38 kcal/mol. It is observable that the three zinc finger 1AAY has a slightly stronger binding than the dimer PUT3 within margin of error (Table 3.11, Table 3.12).

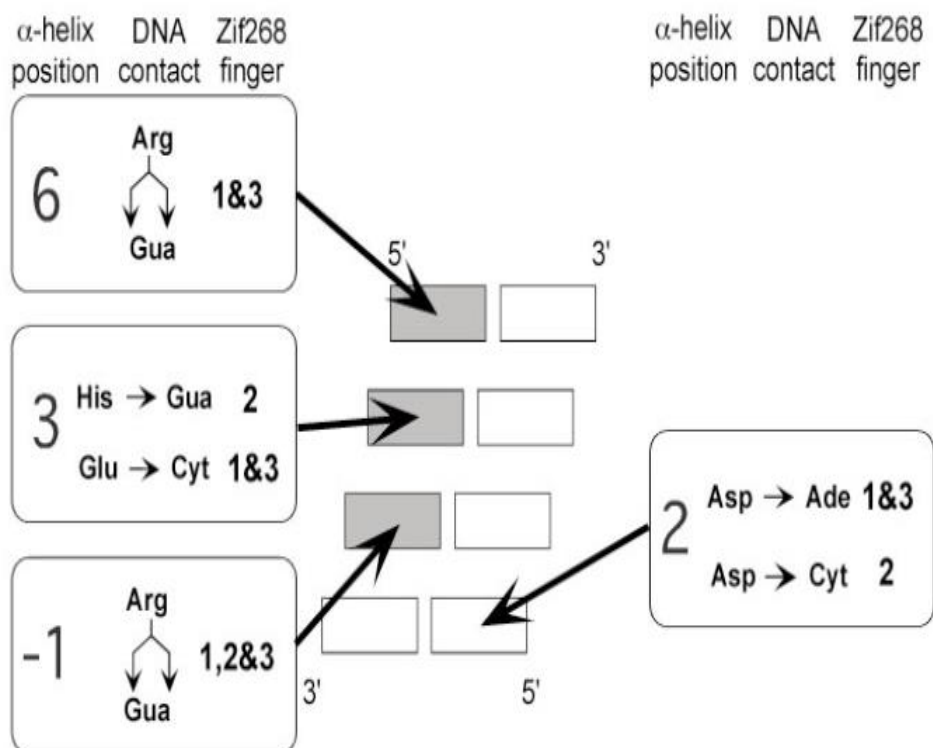


Figure 3.20: The base contact of the individual fingers of 1AAY-DNA complex. Individual base are represented by rectangles, the primary strand of the DNA in the left and the secondary strand on the right. Shaded rectangles represents the 3 bases in the primary strand that are contacted by each finger<sup>12</sup>.

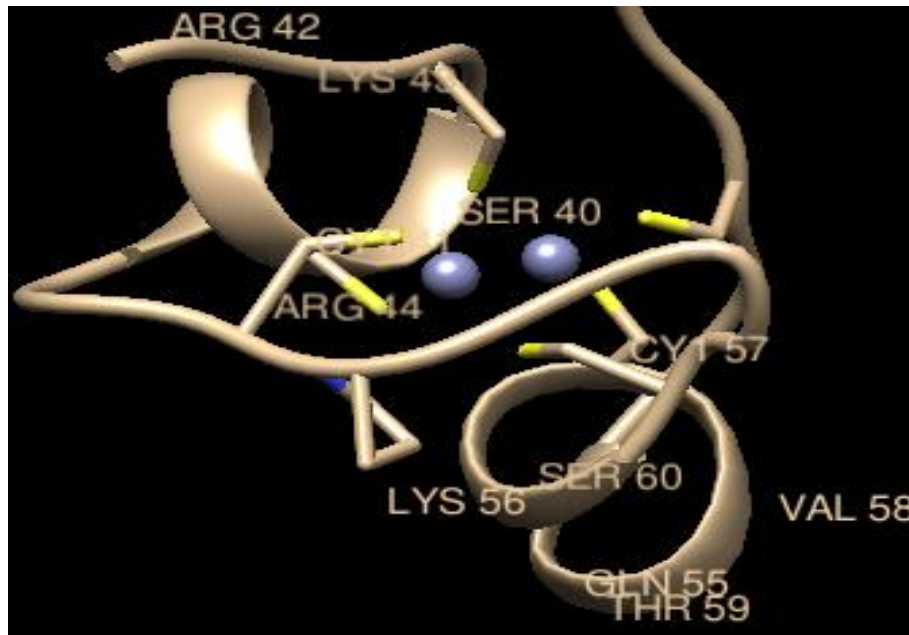


Figure 3.21: The Binuclear cluster region of PUT3 after molecular dynamic simulations showing the residues at position 42 (Arg), and 44 (Arg) in both nuclear clusters are responsible for the contacts to the CGG end of the DNA.

### 3.1.3.3 Comparison of PUT3-DNA Complex with PPR1-DNA Complex (Binuclear Cluster)

Both complexes PUT3-DNA, and 1PYI-DNA showed a conserved set of contacts with the CGG site near each end. Both proteins have a special component which is half site recognition in this component only one amino acid side chain is used to contact CGG DNA site near each end. In PPR1 this side chain is lysine but in PUT3 it is mostly a histidine residue where the nitrogen atom in side chain with both proteins is used to make hydrogen bonds<sup>38,44</sup>.

The major difference between these two proteins is in the arrangements of the dimerization and linker domains and in how these arrangements render to specify

DNA site separation. In 1PYI the linker folds in  $\beta$ -hairpins which form the basis for the helices of the coiled coil and the N termini of these helices. This is stabilized by the Zn domains. This creates a tight and homogeneous structure with a short DNA binding site with a 6 base pairs that are separated the CGG half sites (Figure 3.12). But in the PUT3 protein a 10 base pairs separate the CGG half site. These residues are specified through the minor groove DNA interactions due to the unique arrangement of the linker and dimerization element of the PUT3<sup>38,44</sup>(Figure 3.19).

#### **3.1.3.4 Binding Affinities of PUT3 protein to DNA and its Mutants to DNA**

PUT3 binding energy calculated by MM/GBSA is close to the experimental value  $\Delta G_{\text{exp}} = -11.20$  kcal/mol, while that calculated using MM/PBSA was +2.1 kcal/mole. The latter value is far from the experimentally found values (Table 3.11, Table 3.12).

In vitro binding studies of PUT3 show that PUT3 protein is sensitive to mutation of the base pair separating DNA half site<sup>44</sup>. For that reason we performed a mutation in the separation region at the complementary chain (3`-5`) where the T base nucleotides at positions 24 and 30 are replaced by 5IU unit. (Figure 3.22 and Figure 3.23).

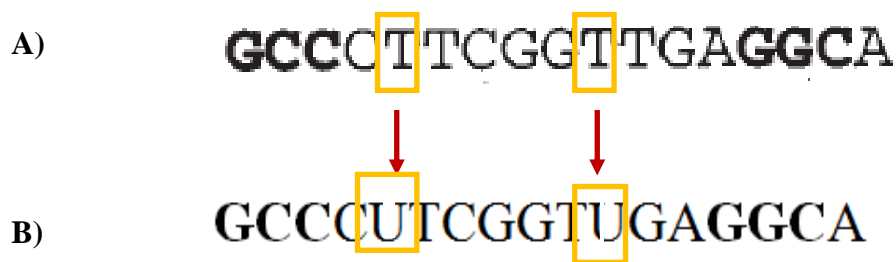


Figure 3.22: A) Complementary chain (3'-5') of the PUT3-DNA complex. B) Mutant complementary chain (3'-5') of the PUT3-DNA complex.

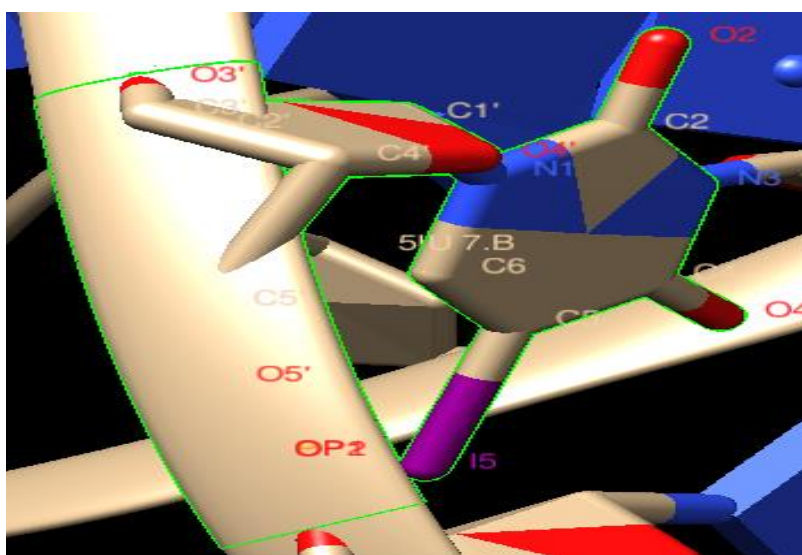


Figure 3.23: Schematic representation of the 5IU unit.

PUT3 mutant (T24U and T30U) gave a positive binding energy as calculated by both the MM/GBSA and MM/PBSA methods with a  $\Delta\Delta G_{\text{MM/GBSA}}$  absolute +16.39 kcal/mole relative to the wild type PUT3 binding (Table 3.11). This result indicates

that the reaction favored the dissociation rather than the association<sup>8</sup>. The PUT3 has a higher specificity and affinity to the wild type DNA and the mutation in the separation region proved to be very sensitive to binding<sup>44</sup>.

Table 3.11. Calculated energies (kcal/mol) for 1AAY and 1ZME DNA complexes: Binding free energy ( $\Delta H$ ) was calculated in a water box and using MM/GBSA and MM/PBSA, the net energy ( $\Delta G$ ) was calculated by subtracting  $T\Delta S$  from *nmode* calculations from the binding free energy.

Complexes	$\Delta G_{\text{binding}}$ ( $\Delta H$ ) (MM/GBSA) kcal/mol	$\Delta G_{\text{binding}}$ ( $\Delta H$ ) (MM/PBSA) kcal/mol	$T\Delta S$ Kcal/mol	$\Delta G_{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G_{\text{binding}}-T\Delta S$	$\Delta G_{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G_{\text{binding}}-T\Delta S$
<b>1AAY</b>  (3F)	-84.09	-106.7	-88.36	+4.270	-18.38
<b>1ZME</b>  (Binuclear cluster)	-99.25	-81.36	-83.46	-15.79	+2.100
<b>1ZME mutant</b>  (Binuclear cluster)	-92.22	-80.98	-92.82	+0.60	+12.44

Table 3.12. Comparison of theoretically calculated  $k_d$  values from both MM/GBSA and MM/PBSA with the experimentally derived values using Gel-shift assays.

Complexes	Calculated				Experimental	
	$\Delta G_{\text{absolute}}$ MM/GBSA kcal/mol $=\Delta G_{\text{binding}}-T\Delta S$	$\Delta G_{\text{absolute}}$ MM/PBSA kcal/mol $=\Delta G_{\text{binding}}-T\Delta S$	$k_d$ $=10^{(\Delta G_{\text{MM/GBSA}}/1.4)}$ nM	$k_d$ $=10^{(\Delta G_{\text{MM/PBSA}}/1.4)}$ nM	$k_d$ nM	$\Delta G$ $=1.4\log K_d(M)$ Kcal/mol
<b>1AAY</b>  (3F)	+4.270	-18.38	$11.2 \cdot 10^2$	$7.4 \cdot 10^{-5}$	13	-11.04
<b>1ZME</b>  Binuclear cluster	-15.79	+2.100	$5.26 \cdot 10^3$	$3.1 \cdot 10^{10}$	10	-11.20

### 3.1.3.5 PUT3- DNA Interaction: Hydrogen-Bonding

Each amino-terminal Zn domain binds identically a 10 base pair recognition DNA site with the specific conserved CGG triplets symmetrically placed near each end as a nonsymmetrical homo-dimer in the major groove<sup>44</sup>. Some of the residues in the linker region contact the bases in the minor groove region.

The hydrogen bonds estimated by molecular dynamic simulations were similar to the experimentally found bonds with the exception of the hydrogen bonds between Arg<sub>42</sub> and DC<sub>15</sub>, Arg<sub>142</sub> and DC<sub>32</sub>, His<sub>115</sub> and DG<sub>20</sub> which were missing.

Table 3.13. The direct contacts between binuclear clusters and the 5'-3' sequence and its 3'-5' complementary sequence. The numbering of protein sequence differ by 4 numbers than the X-ray structural (R40 = R44 in our case).

Acceptor		Donor	Distance Å	PYMOL Screen Shot
Zn <sub>2</sub> CYS <sub>6</sub> binuclear cluster 1	Arg <sub>44</sub> -O	DC <sub>2</sub> -H <sub>42</sub>	1.8	
		5'-3'		



	His <sub>45</sub> -ND <sub>1</sub>	DC <sub>32</sub> -H <sub>42</sub> 3'-5'	2.1	
	His <sub>45</sub> -O	DC <sub>33</sub> -H <sub>42</sub> 3'-5'	1.9	
Zn <sub>2</sub> CYS <sub>6</sub> binuclear cluster 2	Arg <sub>144</sub> -O	DC <sub>19</sub> -H <sub>42</sub> 3'-5'	3.8	

	His <sub>145</sub> -O 5'-3'	DC <sub>16</sub> -H <sub>42</sub> 5'-3'	2.2	
Linker	DA <sub>12</sub> -N <sub>3</sub> 5'-3'	Lys <sub>141</sub> -HZ <sub>2</sub>	2.1	
	Lys <sub>142</sub> -O 3'-5'	DG <sub>26</sub> -H <sub>22</sub>	2.7	
	Val <sub>144</sub> -O 3'-5'	DG <sub>27</sub> -H <sub>22</sub>	2.3	
	DG <sub>27</sub> -N <sub>2</sub> 3'-5'	Val <sub>144</sub> -H	3.0	

## **3.2 The Binding Free Energy of Zinc Finger-DNA Complexes: MM/GBSA Versus MM/PBSA**

MM/GBSA and MM/PBSA methods are frequently used in the estimation of the free energy of binding of ligands to biological macro-molecules<sup>71,72</sup>. Theoretically the PB model is more accurate than the GB model, MM/PBSA is superior in the prediction of the binding free energy than MM/GBSA. Many studies were performed to compare these two methods. Some of these studies indicate that MM/PBSA gives better results that agree with those obtained by experimental methods while other researchers disagree with this opinion. A third opinion suggested that both methods have a similar reliability depending on the system under study<sup>71,72</sup>.

In this study the MM/GBSA method gives results that agree well with the experimentally obtained values within error in most cases. In general, MM/PBSA gave better energy prediction upon mutation.

The performance of the MM/GBSA and MM/PBSA methods were compared based on the calculation of the binding affinities of the different protein-DNA complexes (Table 3.11). MM/PBSA showed better correlation with experimentally obtained energies ( $R = 0.59$ ) than MM/GBSA which gave a correlation coefficient value of 0.69, but both methods gave poor correlation with experimental values.

Theoretically, MM/PBSA is more accurate than MM/GBSA because the PARSE radii are on average smaller than the modified Bondi radii so the dielectric boundary is closer to the charge center in PB leading to a lower binding free energy<sup>73</sup>.

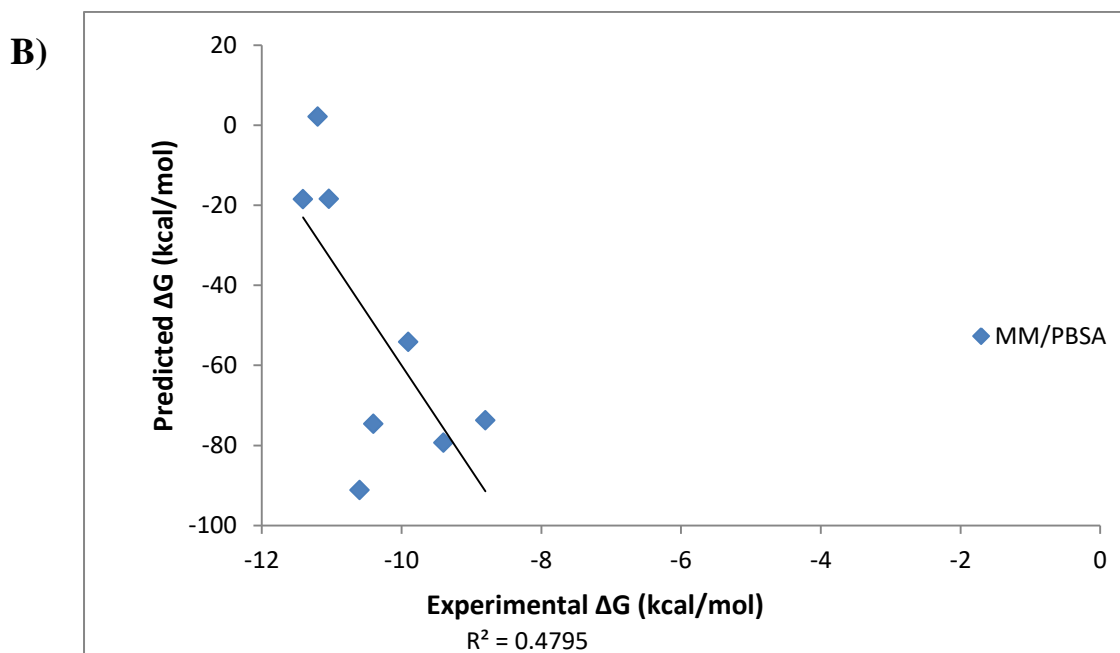
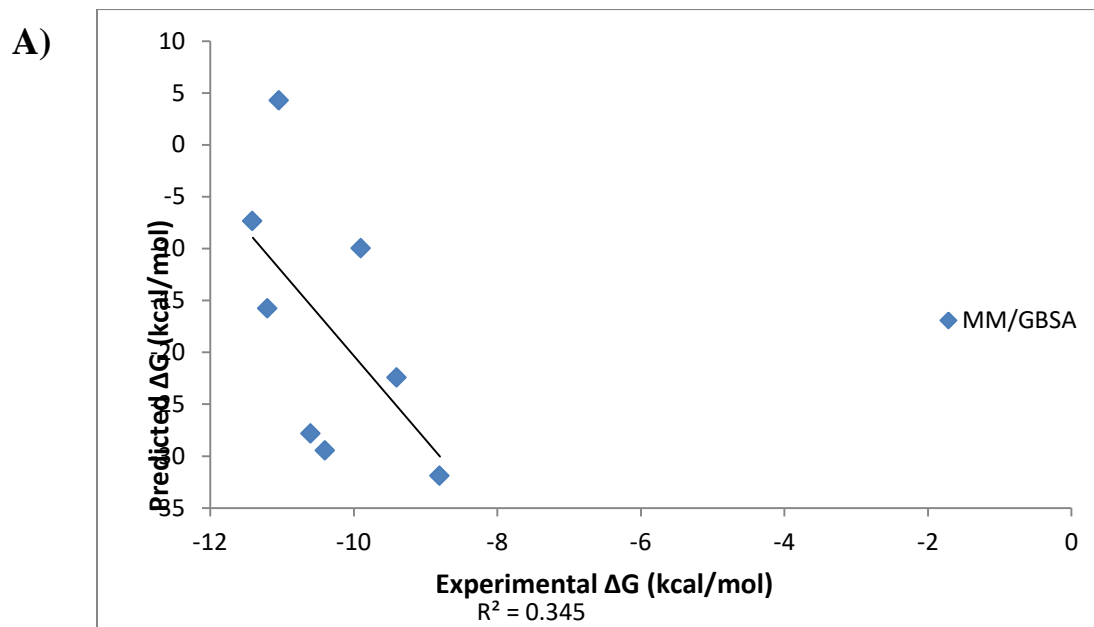


Figure 3.24: The correlation between the binding free energies calculated by A) MM/GBSA, B) MM/PBSA methods and the experimental values.

Table 3.14: Calculated and experimental binding free energies (Kcal/mol) using MM/GBSA and MM/PBSA methods for our protein-DNA complexes.  $G_{ELE}$  is multiplied by  $10^2$ .

<b>Complexes</b>	<b>VdW (kcal/mol)</b>	<b><math>G_{ELE} * 10^2</math> (kcal/mol)</b>	<b><math>\Delta G_{MM/GBSA}</math> (kcal/mol)</b>	<b><math>\Delta G_{MM/PBSA}</math> (kcal/mol)</b>	<b><math>\Delta G_{Experimental}</math> (kcal/mol)</b>
<b>1AAV</b>	-110.1	-51.91	4.270	-18.38	-11.04
<b>5EGB</b>	-175.4	-111.5	29.48-	-74.66	-10.40
<b>L9/24</b>	-173.8	-98.39	31.92-	-73.77	-8.829
<b>L13</b>	-175.4	-118.7	27.83-	-91.19	-10.60
<b>L20<sub>bindAT</sub></b>	-172.8	-111.4	22.44-	-79.33	-9.346
<b>L20<sub>bindGC</sub></b>	-172.3	-111.8	9.970-	-54.20	-9.857
<b>1PYI</b>	-84.14	-44.61	-7.350	-18.52	-11.41
<b>1ZME</b>	-139.9	-50.50	-15.79	+2.100	-11.20

## Chapter 4

# CONCLUSION

The affinity of multiple fingers and ZnF dimmers showed different behavior and a binding modes to that reported for three Znf proteins. PRDM9<sub>A</sub>, a four zinc finger protein and its variant alleles make contacts to the DNA major groove where most of the hydrogen bond contacts to purine bases (G or A) in the complementary strand (3`-5`) and residues at position -2, -1, 3, 6 vary in their degree of contact to the DNA according to the type of finger involved in binding. PPR1 binds to a 14 base pair recognition DNA site with conserved specificity to CGG triplets symmetrically placed near each end as a nonsymmetrical homo-dimer through major groove. None of the residues in the linker region and the coiled-coil contact the base nucleotide. Instead, they make contacts with the sugar phosphate backbone only. PUT3 binds a 10 base pair recognition DNA site to the specific conserved CGG triplets symmetrically placed near each end as a nonsymmetrical homo-dimer through major groove and some of the residues in the linker region contact the base nucleotide in the minor groove region. The  $\alpha$ -helix Zif268 bind the major groove of the target DNA through the surface amino acids side chains at position -1, 2, 3 and 6. Residues at positions -1, 3, 6 bind to three bases of the

primary strand. The residue at position 2 binds the fourth base in the complementary strand (5`-3`).

The binding free energies of different protein-DNA complexes reflect a change in affinity and specificity. The calculated absolute binding free energy of PRDM9<sub>A</sub> (4F) ( $\Delta G_{MM/GBSA} = -29.48$  kcal/mol, and  $\Delta G_{MM/PBSA} = -74.66$  kcal/mol) showed a higher affinity and specificity than Zif268 ( $\Delta G_{MM/GBSA} = +4.27$  kcal/mol, and  $\Delta G_{MM/PBSA} = -18.38$  kcal/mol) and  $\Delta G_{exp} = -11.04$  kcal/mol . PPR1 dimer ( $\Delta G_{MM/GBSA} = -7.35$  kcal/mol, and  $\Delta G_{MM/PBSA} = -18.52$  kcal/mol) and PUT3 dimer the absolute energy ( $\Delta G_{MM/GBSA} = -15.79$  kcal/mol, and  $\Delta G_{MM/PBSA} = +2.1$  kcal/mol) was almost the same as for Zif268 (3F) which indicates a similar affinity and specificity to the target DNA.

Pabo<sup>12</sup> suggested that to tune the specificity and affinity of ZnF to DNA longer linkers are needed and in other cases dimerization is desired. But the increase of the number of fingers results in a plateau of the binding constant values for three fingers and above. Due to the different mode of interaction in the four fingers PRDM9<sub>A</sub>,  $\Delta G$  was calculated to be -56.28 kcal/mol which is higher than that for three finger proteins and from values calculated by Pabo<sup>12</sup>.

The calculated absolute free energy for different complexes were very close to those found by experiment within the error. The entropy changes in different

complexes are unfavorable (negative values) which confirms that this process is driven by ( $\Delta H$ ). The changes in values of absolute energy  $\Delta G$  are parallel to the changes in the binding energy ( $\Delta H$ ) in when both MM/GBSA and MM/PBSA were used.

$\Delta H$  showed higher correlation with contributing energies in MM/GBSA (to non-polar and van der Waals energies with R values 0.97). While in MM/PBSA the correlation is higher with electrostatic and polar energies with R values 0.95. This difference is due to the way each method handles the calculations.

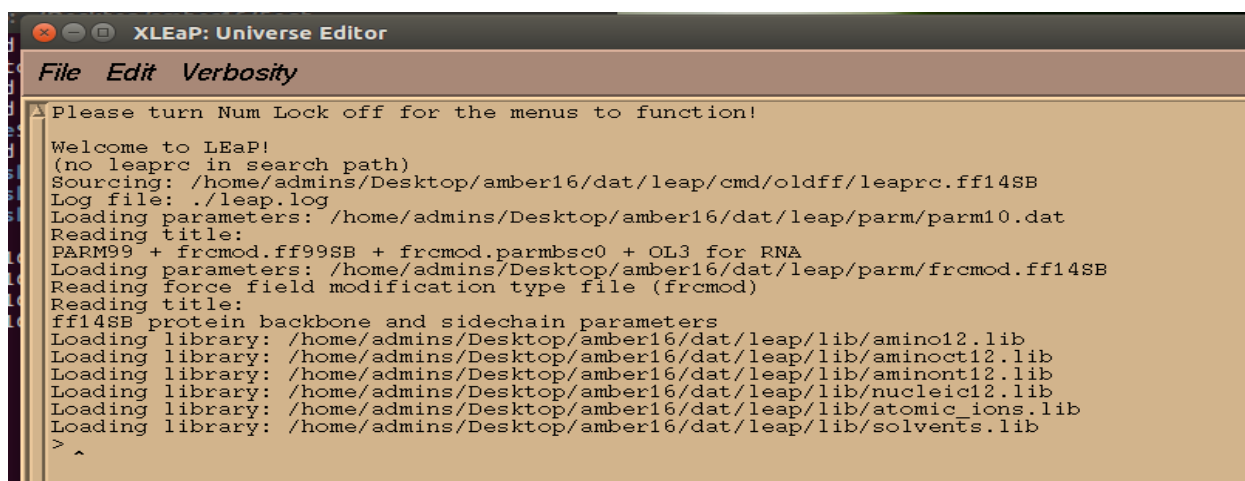
MM/PBSA showed better correlation with experimentally obtained energies than MM/GBSA which gave a lower correlation coefficient value. MM/GBSA gives results that agree better with the experimentally obtained values within error for most of complexes. In general, MM/PBSA gave better energy prediction upon mutation and upon changing of residues in the protein.



## APPENDICES

**Appendix A:** a) preparing the *Prmtop* and *Inpcrd* files for AMBER using *Xleap* Editor, and b) preparing mutants of protein-DNA complexes.

a) *Prmtop* and *Inpcrd* files refer to the molecular topology/parameter and coordinate files, respectively<sup>74</sup>. These files were created using Xleap editor as shown in Figure A.1:



```
XLEaP: Universe Editor
File Edit Verbosity
Please turn Num Lock off for the menus to function!
Welcome to LEaP!
(no leaprc in search path)
Sourcing: /home/admins/Desktop/amber16/dat/leap/cmd/oldff/leaprc.ff14SB
Log file: ./leap.log
Loading parameters: /home/admins/Desktop/amber16/dat/leap/parm/parm10.dat
Reading title:
PARM99 + frcmod.ff99SB + frcmod.parmbsc0 + OL3 for RNA
Loading parameters: /home/admins/Desktop/amber16/dat/leap/parm/frcmod.ff14SB
Reading force field modification type file (frcmod)
Reading title:
ff14SB protein backbone and sidechain parameters
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/aminol2.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/aminoc12.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/aminont12.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/nucleic12.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/atomic_ions.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/solvents.lib
>
```

Figure A.1: Loading configurations file needed for AMBER ff14SB force field.

The zinc ion files for AMBER PREP, PARMSET and OFF library files were loaded before loading the protein-DNA complex PDB files in Xleap. The commands to load these files are shown in the following screen:

```

> addAtomTypes { {"ZN" "Zn" "sp3"} {"S1" "S" "sp3"} {"N2" "N" "sp3"} }
> loadoff atomic_ions.lib
Loading library: /home/admins/Desktop/amber16/dat/leap/lib/atomic_ions.lib
> loadamberparams frcmod.ions11m_1264_tip3p
Loading parameters: /home/admins/Desktop/amber16/dat/leap/parm/frcmod.ions11m_1264_tip3p
Reading force field modification type file (frcmod)
Reading title:
Li/Merz ion parameters of monovalent ions for TIP3P water model (12-6-4 set)
> loadamberprep ZAFF.prep
Loading Prep file: /home/admins/Desktop/amber16/dat/leap/parm/ZAFF.prep
> loadamberparams ZAFF.frcmod
Loading parameters: /home/admins/Desktop/amber16/dat/leap/parm/ZAFF.frcmod
Reading force field modification type file (frcmod)
Reading title:
ZAFF parameter set created by MTK++/MCPB
>
^

```

Figure A.2: preparing xleap to load the 5EGB, PPR1 and PUT3 PDB files.

In PUT3 complex, the Antechamber program was used to create the force field for residues named 5IU. This program is designed to be used with the "general AMBER force field (GAFF)<sup>75</sup>. The 5IU unit was copied from PUT3 PDB files as a new 5IU.pdb file, and then antechamber command was used to create the "mol2" file:

```
antechamber -i 5IU.pdb -fi pdb -o 5IU.mol2 -fo mol2 -c bcc -s 2
```

This command produced a number of files. These files are all intermediate files used by antechamber and are not required. The most important one is 5IU.mol2 file which defines the needed residue including all charges and atom types (Appendix B).

Then the *Parmchk* command was run to create the parameters files required to load the PDB file in Xleap:

```
parmchk -i 5IU.mol2 -f mol2 -o 5IU.frcmod
```

Then the GAFF force field was loaded to Xleap:

```
source leaprc.gaff
```

5IU.mol2 file and 5IU.frmod were loaded:

```
5IU = loadmol2 5IU.mol2
```

5IU unit was edited in Xleap, the HO<sub>3</sub>' atom that bonded to O<sub>3</sub>' was deleted because at this place its bind to the rest of DNA:

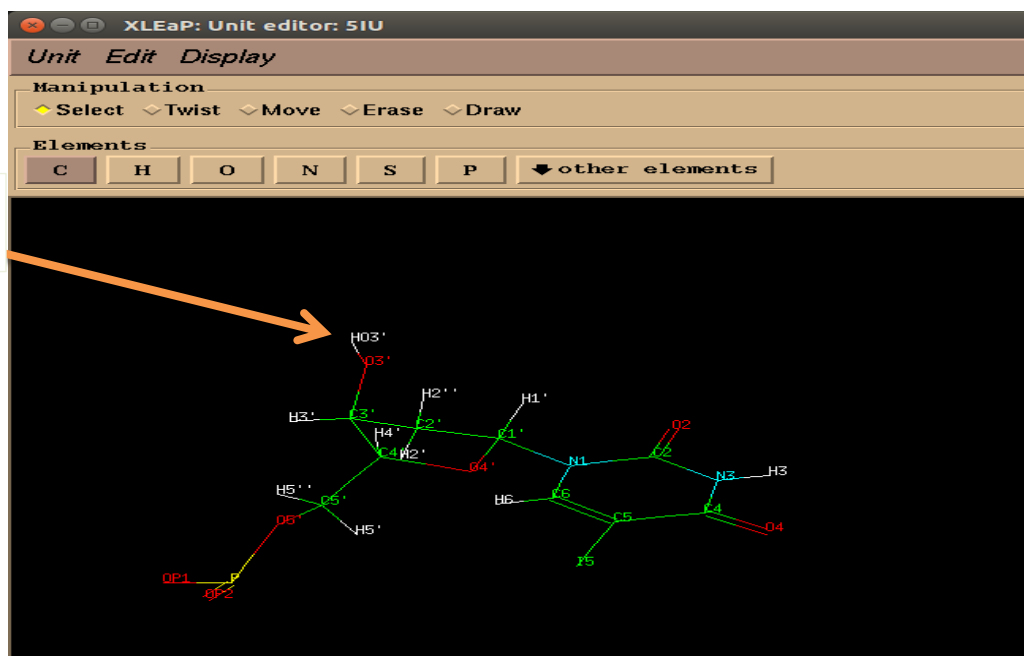


Figure A.3: Graphical representation of 5Iu unit.

Then, the 5IU parameter is saved as a library file and this file is loaded again to ensure the force field of 5IU unit.

```
Saveoff 5IU 5IU.lib
```

Xleap after that was ready to load our three PDB files. For example, 5egb\_dry.pdb file was loaded into xleap by typing in xleap window "e = loadpdb 5egb\_dry.pdb" then the structure of our PDB files were examined using the command check and it's found to be OK.

To bond the zinc ion with the residues in 5EGB, PPR1 and PUT3, respectively commands showed in Appendix C.

Then the *prmtop* and *inpcrd* files for 5egb\_dry.pdb were saved as 5egb\_dry.prmtop and 5egb-dry.inpcrd using the saveamberparm command. This step was repeated for all other protein-DNA complexes. An example shown in edit window of *Xleap* for 5egb\_dry.pdb (Figure A.4).

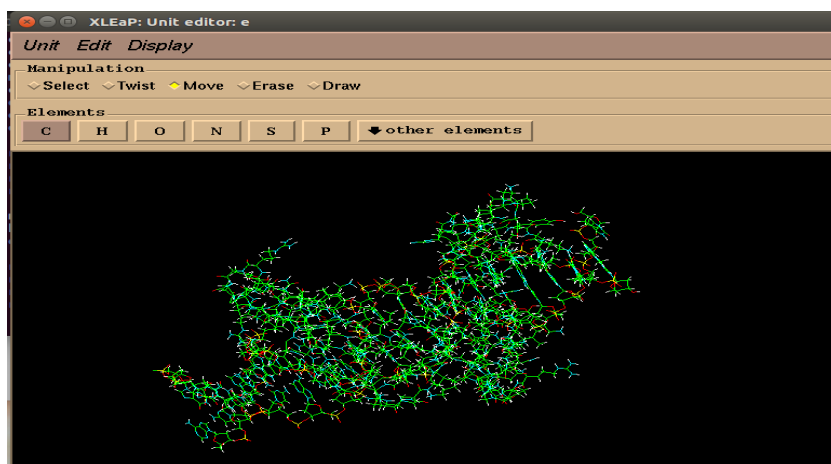
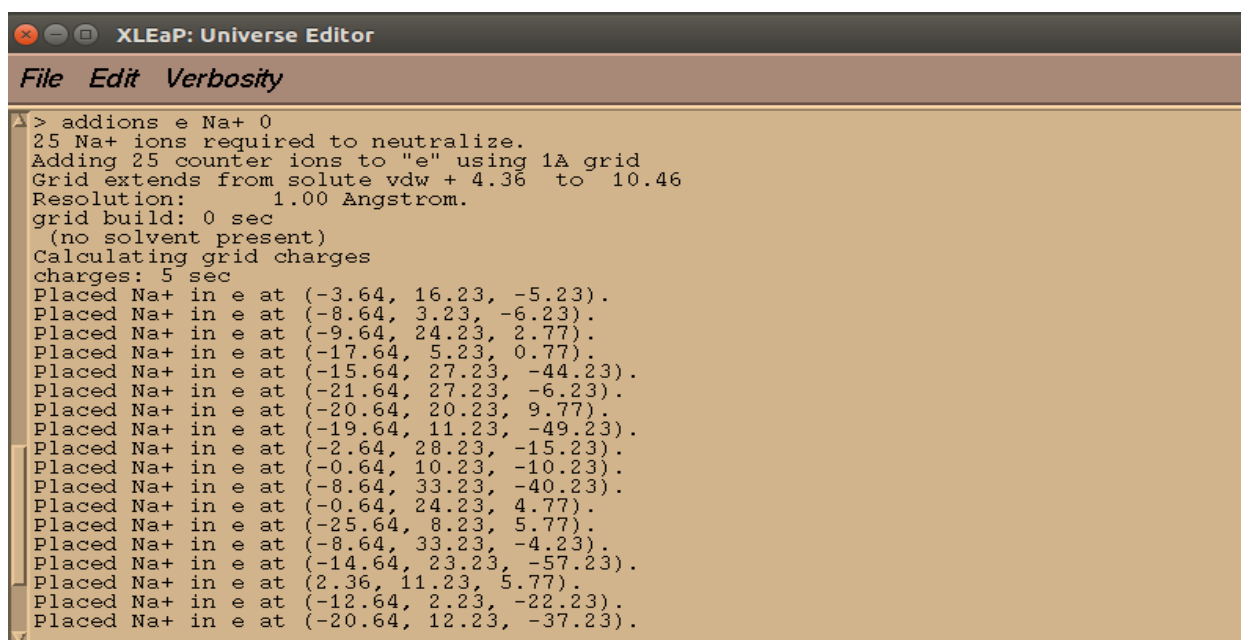


Figure A.4: The editor window of xleap showing the graphical representation of bonded 5egb\_dry.pdb.

Then the system was neutralized by adding sodium ions:

```
addions e Na+ 0
```

This causes a columbic potential on a grid of 1Å resolution and then it puts the counter ions simultaneously at the points of lowest/greatest electrostatic potential (Figure A.5).



```
XLEaP: Universe Editor
File Edit Verbosity
A> addions e Na+ 0
25 Na+ ions required to neutralize.
Adding 25 counter ions to "e" using 1Å grid
Grid extends from solute vdw + 4.36 to 10.46
Resolution: 1.00 Angstrom.
grid build: 0 sec
(no solvent present)
Calculating grid charges
charges: 5 sec
Placed Na+ in e at (-3.64, 16.23, -5.23).
Placed Na+ in e at (-8.64, 3.23, -6.23).
Placed Na+ in e at (-9.64, 24.23, 2.77).
Placed Na+ in e at (-17.64, 5.23, 0.77).
Placed Na+ in e at (-15.64, 27.23, -44.23).
Placed Na+ in e at (-21.64, 27.23, -6.23).
Placed Na+ in e at (-20.64, 20.23, 9.77).
Placed Na+ in e at (-19.64, 11.23, -49.23).
Placed Na+ in e at (-2.64, 28.23, -15.23).
Placed Na+ in e at (-0.64, 10.23, -10.23).
Placed Na+ in e at (-8.64, 33.23, -40.23).
Placed Na+ in e at (-0.64, 24.23, 4.77).
Placed Na+ in e at (-25.64, 8.23, 5.77).
Placed Na+ in e at (-8.64, 33.23, -4.23).
Placed Na+ in e at (-14.64, 23.23, -57.23).
Placed Na+ in e at (2.36, 11.23, 5.77).
Placed Na+ in e at (-12.64, 2.23, -22.23).
Placed Na+ in e at (-20.64, 12.23, -37.23).
```

Figure A.5: neutralization of 5egb complex through the addition of sodium ions.

Final step a 10.0 Å rectangular water box was added (TIP3PBOX) to solvate the complex. (Figure 2.7)

```
solvatebox com TIP3PBOX 10.0
```

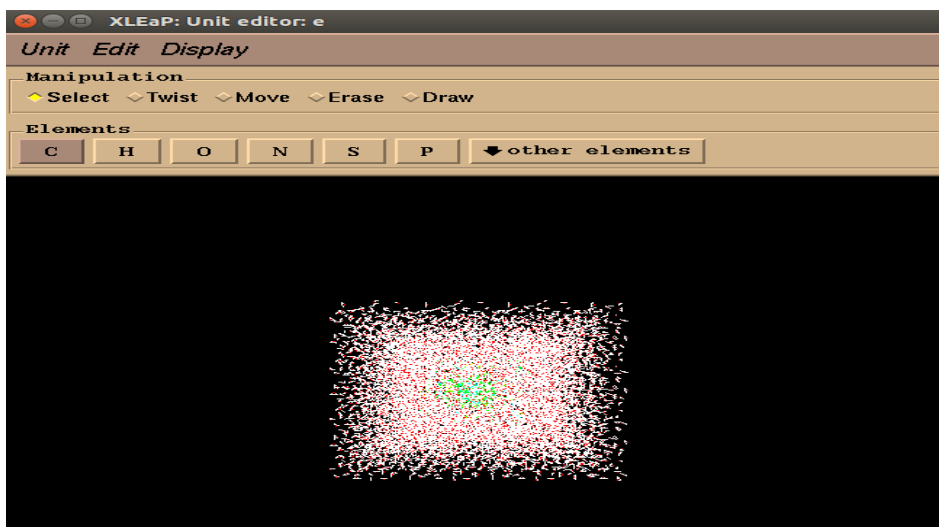


Figure A.6: A graphical representation of 5EGB solvated system.

The *prmtop* and *inpcrd* files for the solvated system were saved:

```
saveamberparm e 5egb_solvated.prmtop 5egb_solvated.inpcrd
```

b) Preparing Mutants of Protein-DNA complexes: (1) mutation of the protein sequence, and (2) mutation of the DNA sequence

Different mutations to protein-DNA complexes were prepared using PyMOL and UCSF Chimera programs<sup>76,77</sup>. PyMOL was used to make mutations to protein sequence. The UCSF Chimera was used to mutate the DNA sequence.

### (1) Mutation of the Protein Sequence

Different single point mutations were introduced using the following steps:

- i. The PRMD9<sub>A</sub>-DNA complex PDB file (5egb\_dry.pdb) was loaded into the *PyMOL*.
- ii. To display the sequence of nucleic bases, amino acids, and zinc ions of the protein-DNA complex the *Display* choice then *sequence* were selected.
- iii. *Mutagenesis* choice was selected from the *Wizard* menu.
- iv. The desired amino acid was chosen from the sequence that showed in the *PyMOL* viewer window.
- v. *No mutation* choice was selected, then the desired amino residue was selected as shown in Figure 2.8.

- vi. The *File* choice, and *save molecule* were selected. Then the file was saved as 5egb\_dry\_L9/24.pdb to indicate mutation for all alleles.

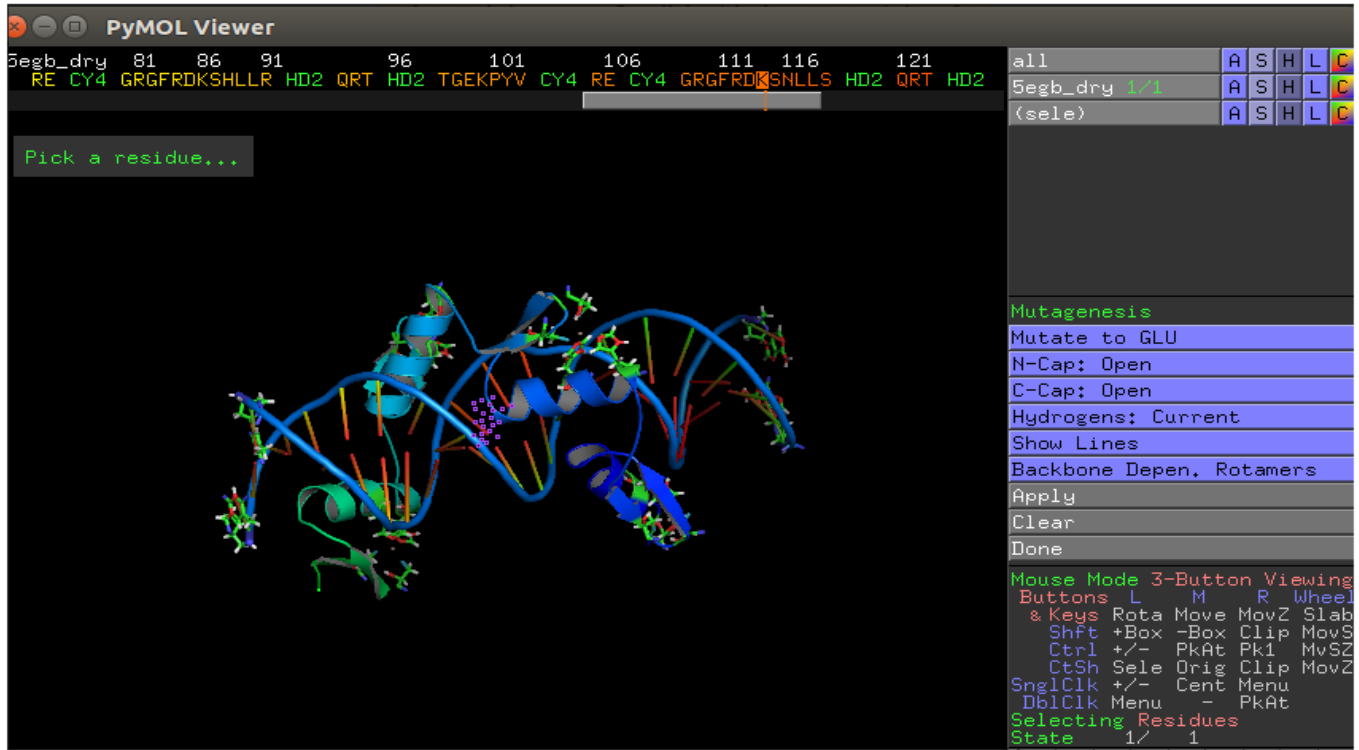


Figure a.7: The PyMOL viewer window showing how the amino Lysine (K) at position 114 in 5EGB-DNA complex is mutated by choosing Glutamic acid from the mutagenesis list after clicking “No mutation” option.

## (2) Mutation of the DNA Sequence

UCSF Chimera program was used to mutate the DNA sequence bound to allele L20 of the PRDM9, PPR1 and PUT3 using the following steps:

- i. The PRDM<sub>9</sub>-DNA complex PDB file was loaded in *UCSF Chimera* “5egb.pdb” that obtained directly from Protein Data Bank.



- ii. To display the sequence of DNA bases, amino acids , and zinc ions of the protein –DNA complex the *preset* choice then *sequence* were selected giving three chain windows (A, C, D); C for the 5`-3` DNA strand, D for 3`-5` DNA strand, and A is for the protein amino acid sequence.
- iii. To perform the mutation of the DNA strands, the *tool* choice then *command line* were selected.
- iv. The desired nucleotide in PRDM<sub>9</sub> allele L20 was chosen from the sequence that showed in the C and D windows.
- v. The thymine nucleotide in the C chain was selected, then in the command window the following command was written:  

Swapna C :11.c

  
[change residue 11 in chain C from DT to cytosine nucleotide].
- vi. The adenine nucleotide in the D chain was selected, then in the command window the following command was written: (Figure 2.9)  

Swapna G : 12.d

  
[change residue 11 in chain D from DA to guanine nucleotide].
- vii. The *File* choice, and *save PDB* were selected. Then the file was saved as 5egb\_dry\_L20CG.pdb to indicate mutation.

viii. The same steps were repeated for PPR1-DNA and PUT3-DNA complexes.

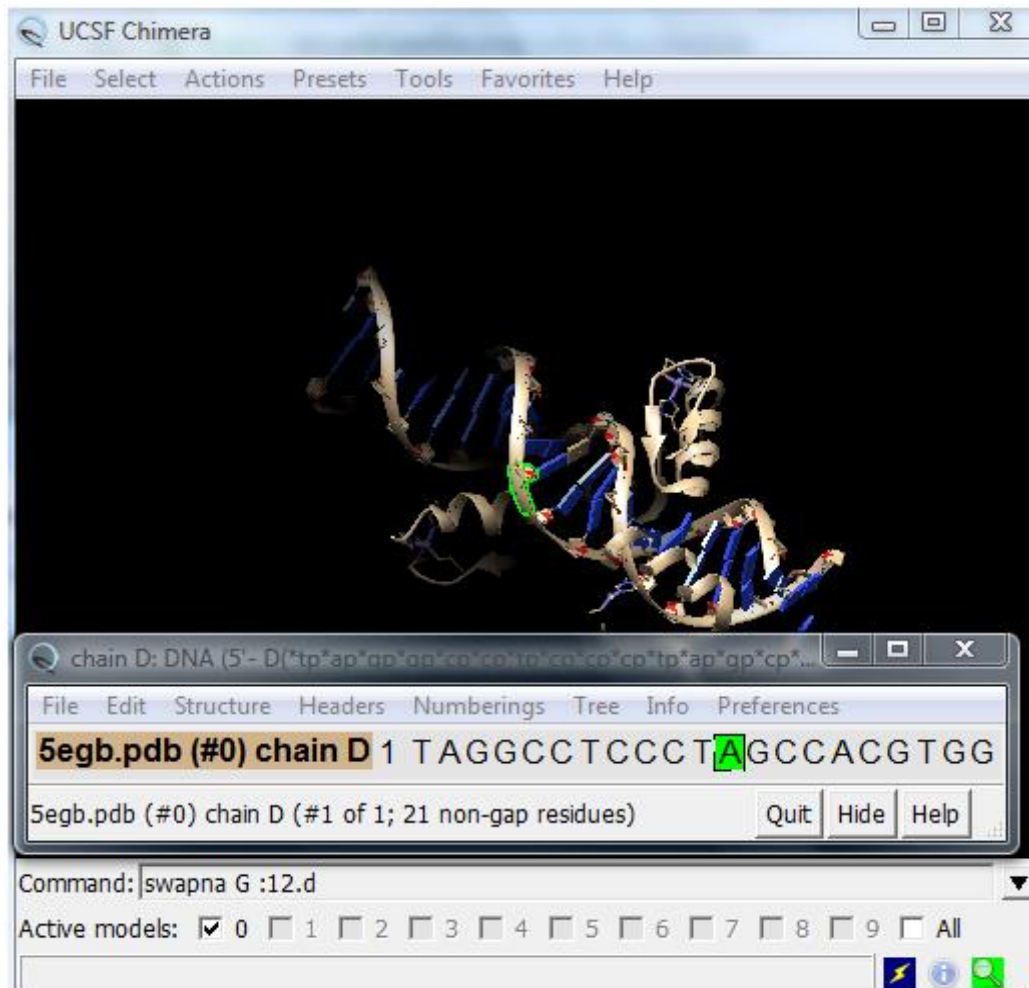


Figure A.8: The UCSF Chimera window showing how the nucleotide adenine (A) at position 12 in chain D in 5EGB-DNA complex is mutated to guanine (G) by written “swapna G :12.d” in the command window.

## Appendix B: 5IU.mol2 file

@<TRIPOS>MOLECULE

u

30 31 1 0 0

SMALL

bcc

@<TRIPOS>ATOM

1 N1	43.7730	3.5510	-9.6420	n	13 u	-0.387000
2 C2	42.4450	3.8220	-9.3870	c	13 u	0.813500
3 N3	41.8340	4.6450	-10.3000	n	13 u	-0.582500
4 C4	42.4090	5.2330	-11.4060	c	13 u	0.729700
5 C5	43.8190	4.9270	-11.5890	cc	13 u	-0.189400
6 C6	44.4190	4.1170	-10.7080	cd	13 u	0.084600
7 O2	41.8540	3.3860	-8.4200	o	13 u	-0.626500
8 O4	41.7290	5.9590	-12.1470	o	13 u	-0.583500
9 I5	44.9510	5.6530	-13.1640	i	13 u	-0.061800
10 C1'	44.4860	2.6090	-8.7560	c3	13 u	0.299500
11 C2'	45.6630	1.8690	-9.3660	c3	13 u	-0.155400
12 C3'	46.3790	1.4060	-8.1110	c3	13 u	0.114100
13 C4'	46.1520	2.5690	-7.1450	c3	13 u	0.105100
14 O3'	45.6310	0.3450	-7.5130	oh	13 u	-0.591800
15 O4'	45.0440	3.3350	-7.6770	os	13 u	-0.405600
16 C5'	47.3390	3.4020	-6.7180	c3	13 u	0.142400
17 O5'	48.2870	3.5100	-7.7710	os	13 u	-0.561200
18 P	49.4260	4.6070	-7.6740	p4	13 u	1.614200
19 OP1	50.2150	4.3650	-6.4400	o	13 u	-0.649000
20 OP2	50.1210	4.7250	-8.9810	o	13 u	-0.697800
21 H5'	47.0400	4.2860	-6.4550	h1	13 u	0.079700
22 H5''	47.7580	3.0010	-5.9400	h1	13 u	0.079700
23 H4'	45.9480	2.1830	-6.2790	h1	13 u	0.100700
24 H3'	47.2960	1.1460	-8.2900	h1	13 u	0.073700

25 H2'	46.2170	2.4460	-9.9150 hc	13 u	0.084200
26 H2''	45.3830	1.1280	-9.9260 hc	13 u	0.084200
27 HO3'	45.6980	-0.3460	-7.9850 ho	13 u	0.415000
28 H1'	43.8090	1.9550	-8.5200 h2	13 u	0.140700
29 H3	41.0010	4.8120	-10.1670 hn	13 u	0.358500
30 H6	45.3220	3.9290	-10.8260 h4	13 u	0.173000

@<TRIPOS>BOND

```

1  1  2 1
2  1  6 1
3  1 10 1
4  2  3 1
5  2  7 2
6  3  4 1
7  3 29 1
8  4  5 1
9  4  8 2
10 5  6 2
11 5  9 1
12 6 30 1
13 10 11 1
14 10 15 1
15 10 28 1
16 11 12 1
17 11 25 1
18 11 26 1
19 12 13 1
20 12 14 1
21 12 24 1
22 13 15 1
23 13 16 1
24 13 23 1
25 14 27 1
26 16 17 1
27 16 21 1
28 16 22 1
29 17 18 1
30 18 19 1
31 18 20 2

```

@<TRIPOS>SUBSTRUCTURE

```

1 u      1 TEMP      0 ***** 0 ROOT

```

**Appendix C:** Commands in Xleap to bond the zinc ion with the residues in 5EGB  
(a), PPR1 (b) and PUT3 (c), respectively.

a- bond e.154.ZN e.48.SG

bond e.154.ZN e.51.SG

bond e.154.ZN e.64.NE2

bond e.154.ZN e.68.NE2

bond e.154.ZN e.76.SG

bond e.154.ZN e.79.SG

bond e.154.ZN e.92.NE2

bond e.154.ZN e.96.NE2

bond e.154.ZN e.104.SG

bond e.154.ZN e.107.SG

bond e.154.ZN e.120.NE2

bond e.154.ZN e.124.NE2

bond e.154.ZN e.132.SG

bond e.154.ZN e.135.SG

bond e.154.ZN e.148.NE2

bond e.154.ZN e.145.NE2

b- bond i.187.ZN i.33.SG

bond i.187.ZN i.50.SG

bond i.187.ZN i.53.SG

bond i.187.ZN i.60.SG

bond i.188.ZN i.33.SG

bond i.188.ZN i.36.SG

bond i.188.ZN i.43.SG

bond i.188.ZN i.50.SG

bond i.189.ZN i.121.SG

bond i.189.ZN i.138.SG

bond i.189.ZN i.141.SG

bond i.189.ZN i.148.SG

bond i.190.ZN i.121.SG

bond i.190.ZN i.124.SG

bond i.190.ZN i.131.SG

bond i.190.ZN i.138.SG

c- bond z.175.ZN z.38.SG

bond z.175.ZN z.41.SG

bond z.175.ZN z.48.SG

bond z.175.ZN z.54.SG

bond z.176.ZN z.38.SG

bond z.176.ZN z.54.SG

bond z.176.ZN z.57.SG

bond z.176.ZN z.64.SG

    bond z.177.ZN z.108.SG

bond z.177.ZN z.111.SG

bond z.178.ZN z.118.SG

bond z.178.ZN z.124.SG

    bond z.175.ZN z.108.SG

bond z.175.ZN z.124.SG

bond z.175.ZN z.127.SG

bond z.175.ZN z.134.SG

## **Appendix D:** Input files for simulations.

### **File 1:** min.in

```
5EGB-Minimization
```

```
&cntrl
```

```
imin=1,maxcyc=1000,ncyc=500,
```

```
cut=8.0,ntb=1,
```

```
ntc=2,ntf=2,
```

```
ntpr=100,
```

```
ntr=1, restraintmask=':1-242',
```

**File 2: heat.in**

(50ps of heating).

```
5EGB - Heat
```

```
&cntrl
```

```
imin=0,irest=0,ntx=1,
```

```
nstlim=25000,dt=0.002,
```

```
ntc=2,ntf=2,
```

```
cut=8.0, ntb=1,
```

```
ntpr=500, ntwx=500,
```



**File 3:** density.in

(50 ps of density equilibration with weak restraints on the complex).

```
5EGB - Density
&cntrl
imin=0,irest=1,ntx=5,
nstlim=25000,dt=0.002,
ntc=2,ntf=2,
cut=8.0, ntb=2, ntp=1, taup=1.0,
ntpr=500, ntwx=500,
ntt=3, gamma_ln=2.0,
temp0=300.0, ig=-1,
```

**File 4:** equil.in

(500ps of constant pressure equilibration at 300K).

```
5EGB - Equil
```

```
&cntrl
```

```
imin=0,irest=1,ntx=5,
```

```
nstlim=250000,dt=0.002,
```

```
ntc=2,ntf=2,
```

```
cut=8.0, ntb=2, ntp=1, taup=2.0,
```

```
ntpr=1000, ntwx=1000,
```

```
ntt=2, gamma_ln=2.0
```

**File 5:** prod.in

(A total of 2 ns production was run and recording the coordinates every 10 ps).

```
5EGB - Prod
&cntrl
imin=0,irest=1,ntx=5,
nstlim=250000,dt=0.002,
ntc=2, ntf=2,
cut=8.0, ntb=2, ntp=1, taup=2.0,
```

**File 6:** calculate-energy.in

(The interaction energy and solvation free energy for the complex, receptor and ligand were calculated and average the results to obtain an estimate of the binding free energy).

Input file for running PB and GB

&general

endframe=50, verbose=1,

# entropy=1,

/

**File 7:** calculate-entropy.in

(The normal modes for the complex, receptor and ligand were calculated, and average the results to obtain an estimate of the binding entropy).

Input file for running entropy calculations using NMode

```
&general
```

```
  endframe=50, keep_files=2,
```

```
/
```

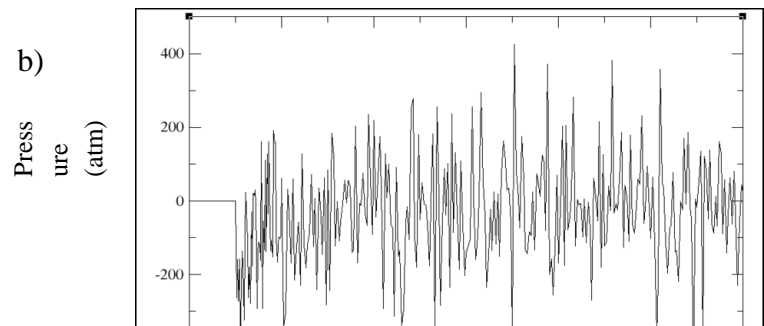
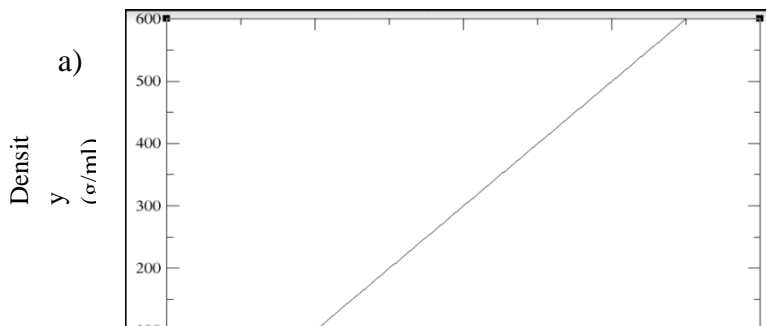
```
&nmode
```

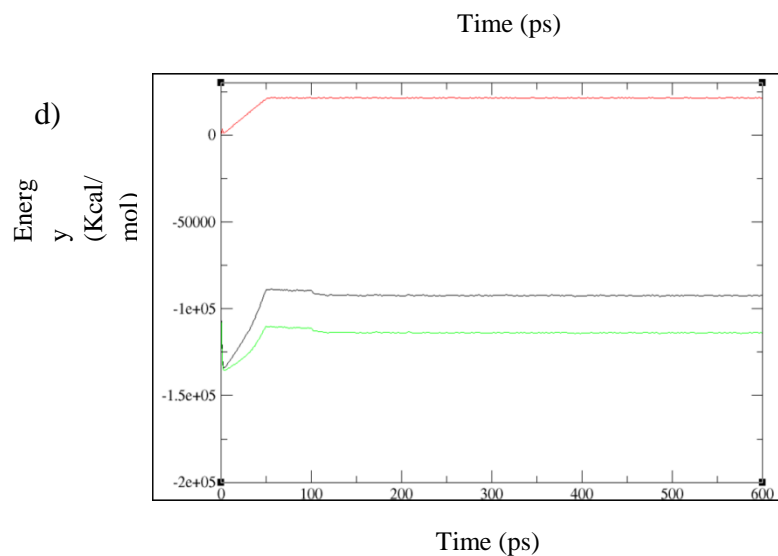
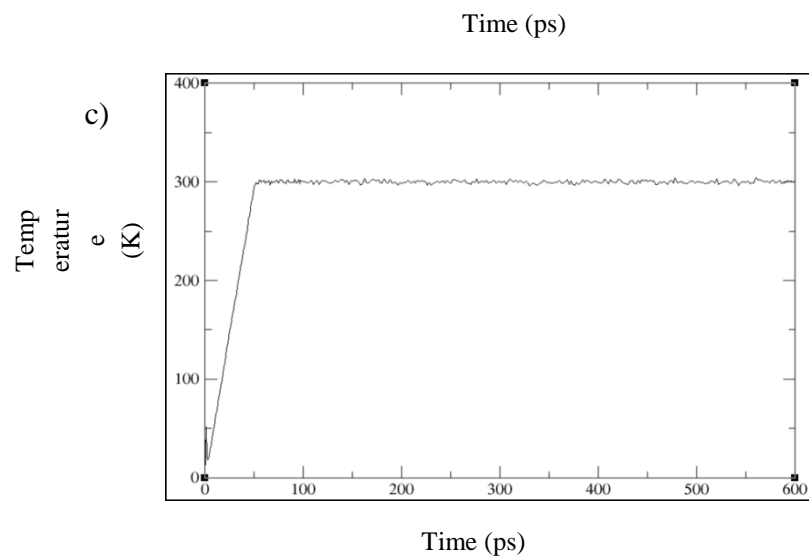
```
  nmstartframe=5, nmendframe=45,
```

```
  nminterval=5, nmode_igb=1, nmode_istrng=0.1,
```

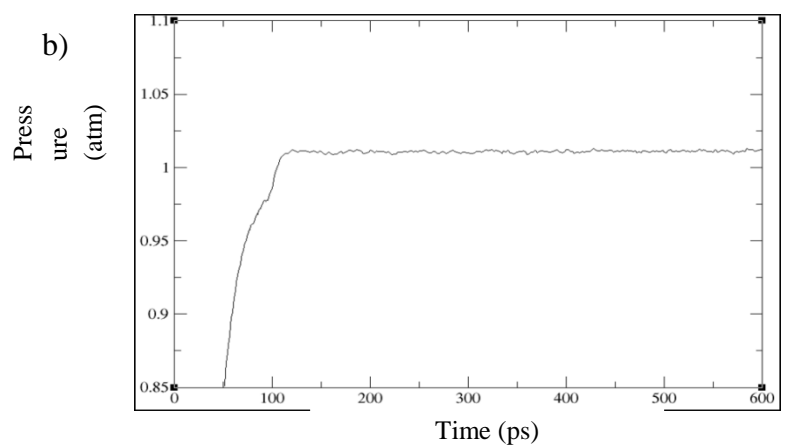
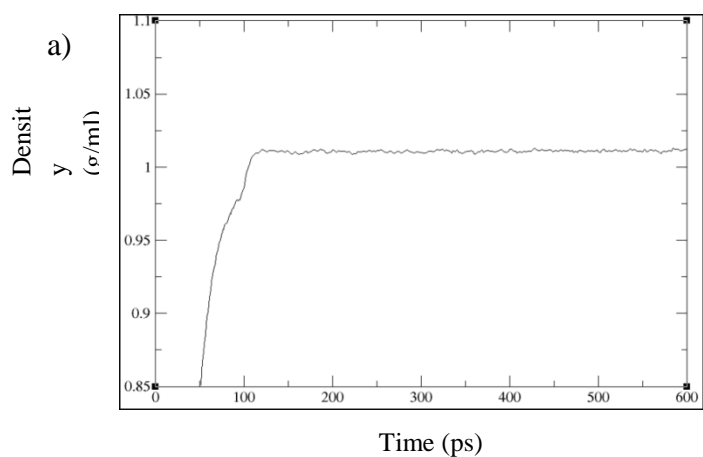
**Appendix E:** Simulation output file for 1PYI (i), and 1ZME (ii), respectively.

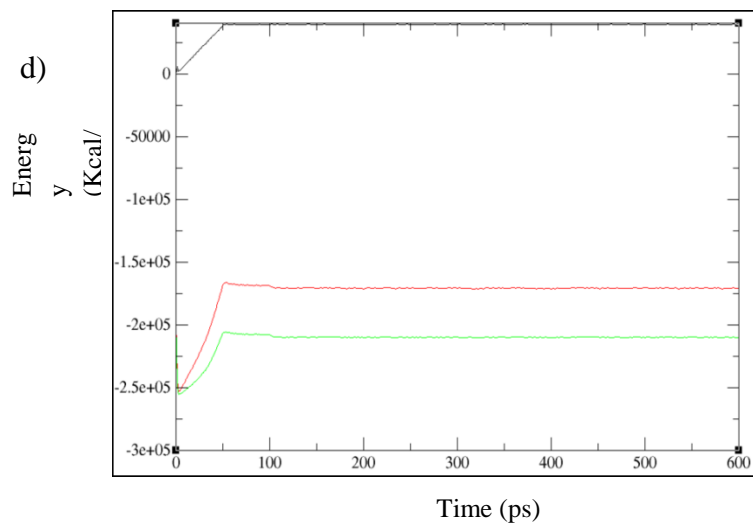
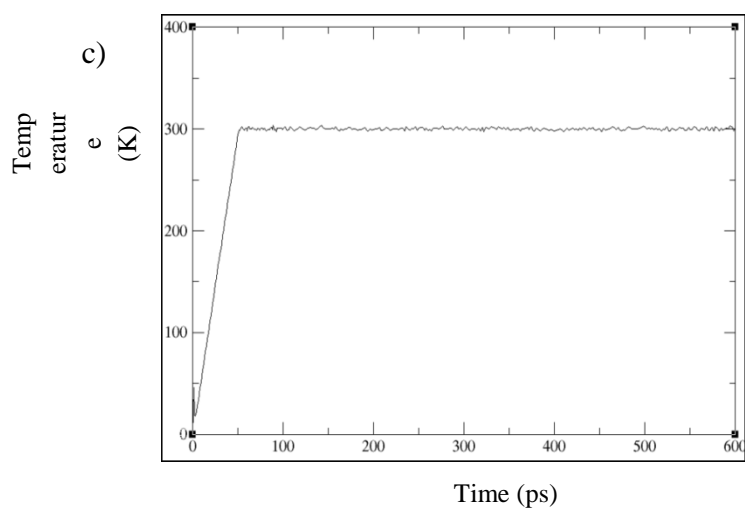
i- 1PYI simulation output plots against time for the heating and equilibration phases of the density (a), temperature (b), pressure (c), and energy (d).





ii- 1ZME simulation output plots against time for the heating and equilibration phases of the density (a), temperature (b), pressure (c), and energy (d).



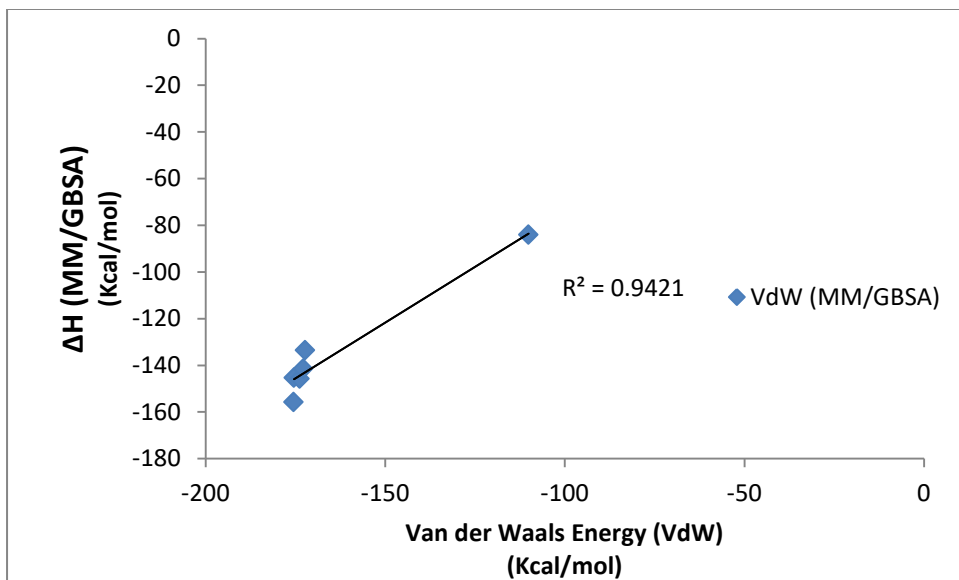


**Appendix F:** Correlation graphs between contribution energy calculated by MM/GBSA (i) & MM/PBSA (ii) and correlation graphs between calculated and experimental dissociation constant ( $K_d$ ) (iii).

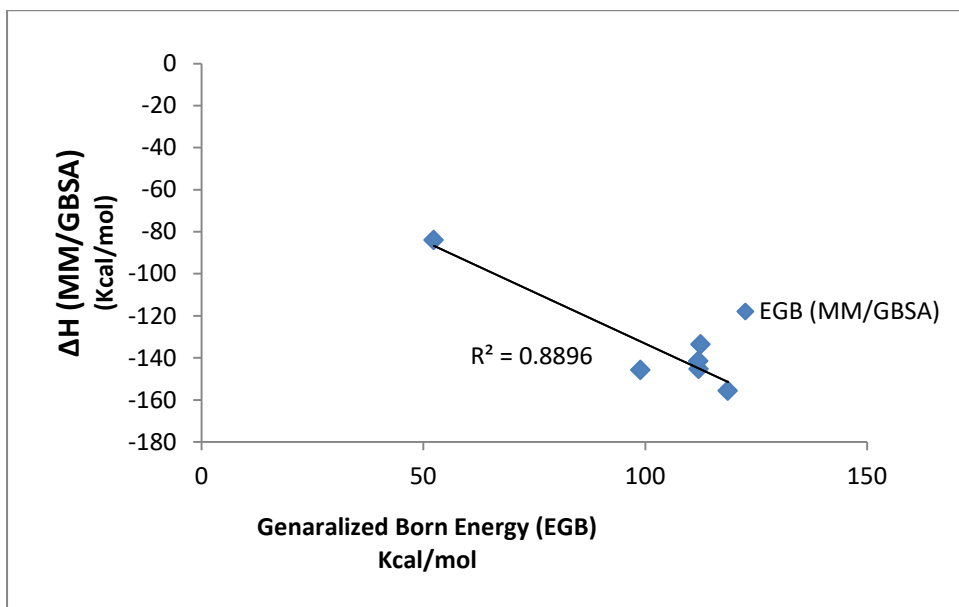
i- Correlation graphs between binding energy calculated by MM/GBSA and van de Waals ( $VdW$ ) (a), electrostatic contribution to the solvation free energy calculated by  $GB(E_{GB})$  (b), non-polar energy ( $E_{SURF}$ ) (C).



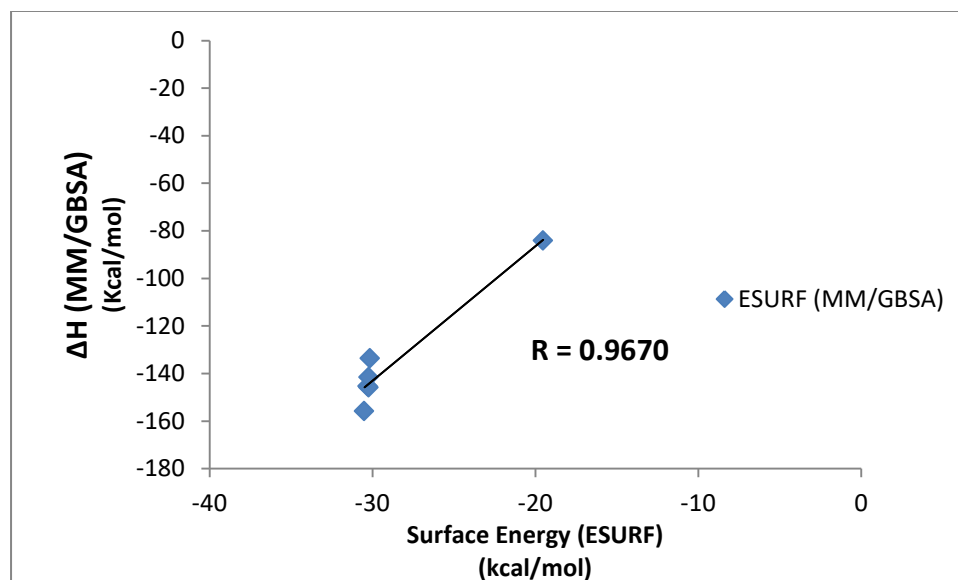
a)



b)

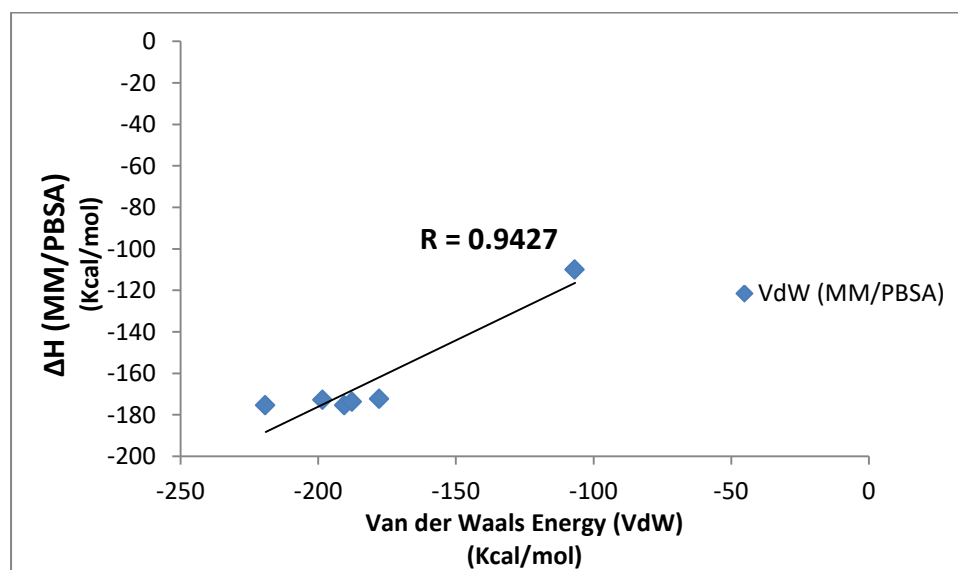


c)

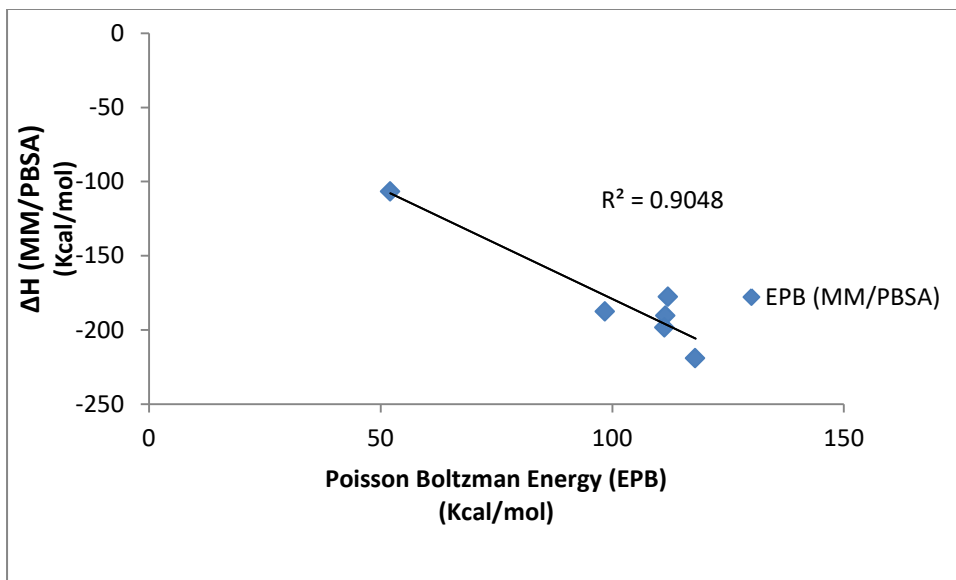


ii- Correlation graphs between binding free energy calculated by MM/PBSA and van der Waals (VdW) (a), electrostatic contribution to the solvation free energy calculated by PB (EPB) (b), non-polar energy (ENPOLAR) (C).

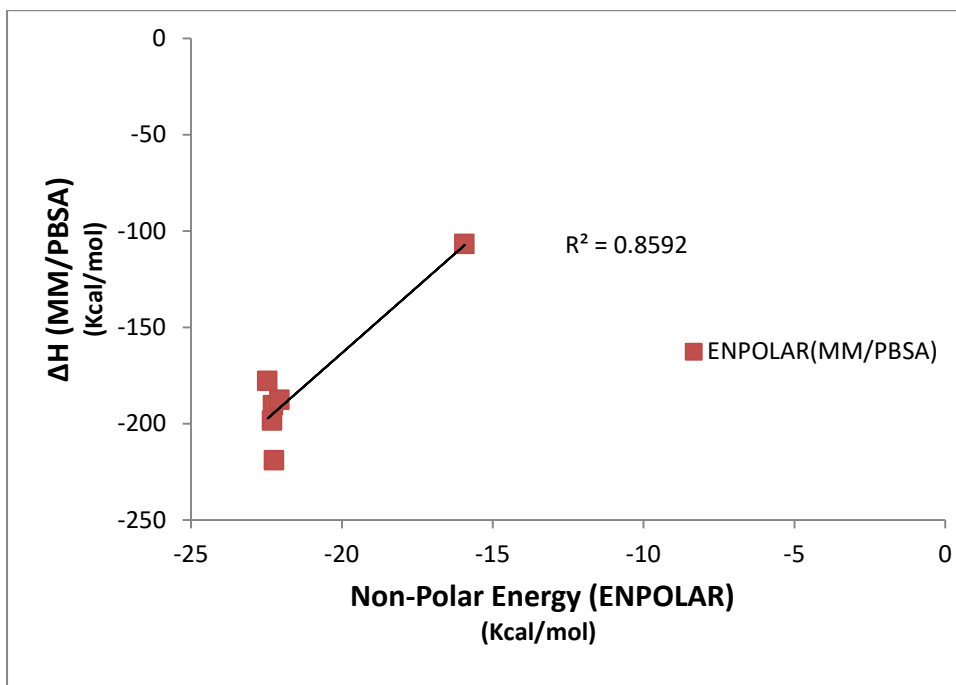
a)



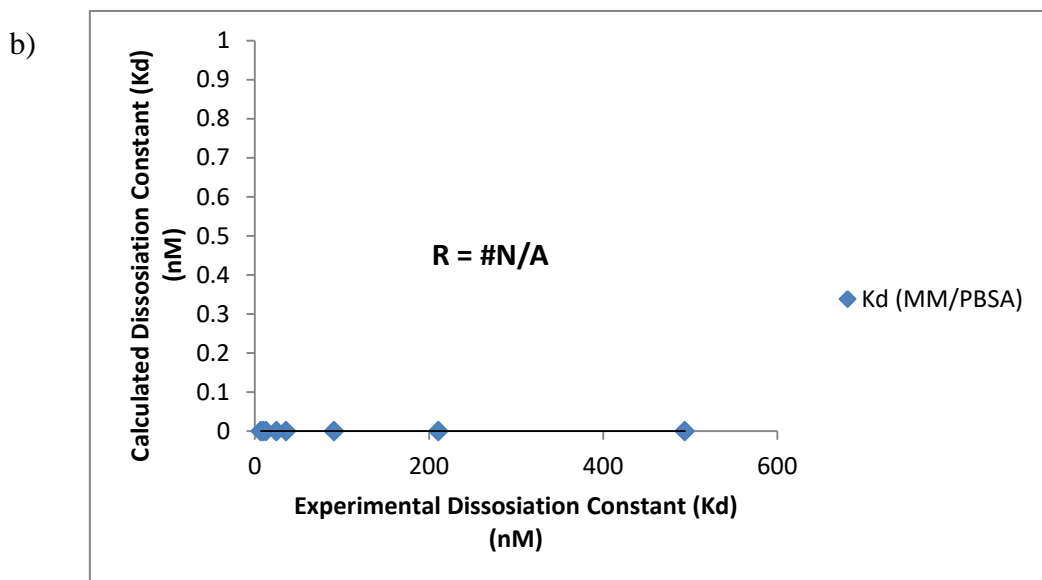
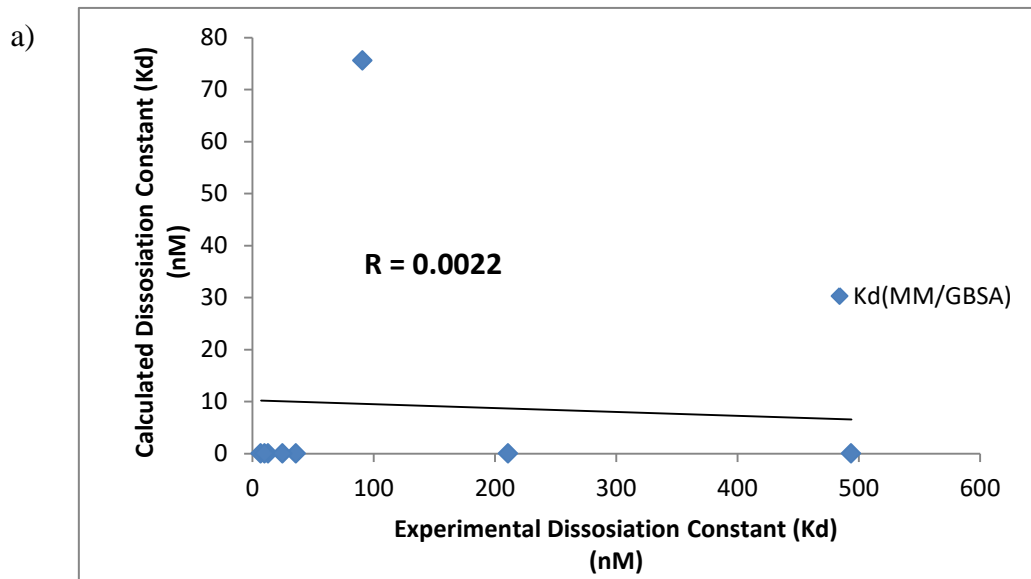
b)



c)



iii- Correlation graphs between calculated and experimental dissociation constant (Kd).



**Appendix J:** Free energy output files calculated using MM/GB(PB)SA for (a) 5EGB, (b) 1PYI, and (c) 1ZME complexes.

(a) Free energy output file for 5EGB.

```
Run on Sat Apr 29 13:49:41 2017
|
|Input file:
|-----
|Input file for running PB and GB
|&general
| endframe=50, verbose=1,
|# entropy=1,
|/
|&gb
| igb=2, saltcon=0.100
|/
|&pb
| istrng=0.100, inp=1, radiopt=0
|/
|-----
|MMPBSA.py Version=14.0
|Solvated complex topology file: 5egb_solvated.prmtop
|Complex topology file:      complex_5egb.prmtop
|Receptor topology file:    receptor_5egb.prmtop
|Ligand topology file:     ligand_5egb.prmtop
|Initial mdcrd(s):         density.mdcrd
|                          equil.mdcrd
|                          heat.mdcrd
|                          prod1.mdcrd
|                          prod2.mdcrd
|                          prod3.mdcrd
|                          prod4.mdcrd
|
|Receptor mask:           ":1-42"
|Ligand mask:             ":43-157"
|
|Calculations performed using 50 complex frames.
|Poisson Boltzmann calculations performed using internal PBSA solver in mmpbsa_py_energy
|
|Generalized Born ESURF calculated using 'LCPO' surface areas
|
|All units are reported in kcal/mole.
|-----
|-----
```

GENERALIZED BORN:

Complex:			
Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1661.3055	11.8265	1.6725

EEL	-4829.0393	29.7433	4.2063
EGB	-7955.1252	16.9784	2.4011
ESURF	82.3043	0.3171	0.0448
G gas	-6490.3447	30.8620	4.3645
G solv	-7872.8210	16.8083	2.3771
TOTAL	-14363.1657	23.4861	3.3214

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-704.9200	7.2024	1.0186
EEL	10261.0620	18.6771	2.6413
EGB	-14871.8884	14.8132	2.0949
ESURF	53.6447	0.1701	0.0241
G gas	9556.1420	19.3461	2.7359
G solv	-14818.2436	14.8430	2.0991
TOTAL	-5262.1016	12.2803	1.7367

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-780.9994	8.5273	1.2059
EEL	-3940.2667	28.4170	4.0188
EGB	-4293.4294	17.7355	2.5082
ESURF	58.9511	0.2376	0.0336
G gas	-4721.2661	27.7275	3.9213
G solv	-4234.4783	17.7249	2.5067
TOTAL	-8955.7444	20.3874	2.8832

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-175.3861	7.6115	1.0764
EEL	-11149.8346	29.4810	4.1692
EGB	11210.1925	27.7014	3.9176
ESURF	-30.2916	0.2230	0.0315
DELTA G gas	-11325.2207	28.9311	4.0915
DELTA G solv	11179.9010	27.6162	3.9055
DELTA TOTAL	-145.3197	5.6453	0.7984

-----  
 -----  
 POISSON BOLTZMANN:

Complex:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
------------------	---------	-----------	-------------------

VDWAALS	-1661.3055	11.8265	1.6725
EEL	-4829.0393	29.7433	4.2063
EPB	-8119.8286	18.4434	2.6083
ENPOLAR	63.6623	0.1475	0.0209
G gas	-6490.3447	30.8620	4.3645
G solv	-8056.1663	18.4074	2.6032
TOTAL	-14546.5111	24.3936	3.4498

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-704.9200	7.2024	1.0186
EEL	10261.0620	18.6771	2.6413
EPB	-15028.6514	15.9017	2.2488
ENPOLAR	40.8436	0.0815	0.0115
G gas	9556.1420	19.3461	2.7359
G solv	-14987.8078	15.9326	2.2532
TOTAL	-5431.6657	12.0197	1.6998

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-780.9994	8.5273	1.2059
EEL	-3940.2667	28.4170	4.0188
EPB	-4248.1688	16.9869	2.4023
ENPOLAR	45.0873	0.1106	0.0156
G gas	-4721.2661	27.7275	3.9213
G solv	-4203.0815	17.0020	2.4044
TOTAL	-8924.3476	19.8046	2.8008

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-175.3861	7.6115	1.0764
EEL	-11149.8346	29.4810	4.1692
EPB	11156.9916	27.8234	3.9348
ENPOLAR	-22.2686	0.1338	0.0189
EDISPER	0.0000	0.0000	0.0000
DELTA G gas	-11325.2207	28.9311	4.0915
DELTA G solv	11134.7230	27.7959	3.9309
DELTA TOTAL	-190.4977	7.2944	1.0316

(b) Free energy output file for 1PYI.

```
| Run on Sat Apr 29 13:50:23 2017
|
| Input file:
|-----
| Input file for running PB and GB
|&general
| endframe=50, verbose=1,
|# entropy=1,
|/
|&gb
| igb=2, saltcon=0.100
|/
|&pb
| istrng=0.100, inp=1, radiopt=0
|/
|-----
|MMPBSA.py Version=14.0
|Solvated complex topology file: 1pyi_solvated.prmtop
|Complex topology file:      complex_1pyi.prmtop
|Receptor topology file:    receptor_1pyi.prmtop
|Ligand topology file:     ligand_1pyi.prmtop
|Initial mdcrd(s):         density.mdcrd
|                           equil.mdcrd
|                           heat.mdcrd
|                           prod1.mdcrd
|                           prod2.mdcrd
|                           prod3.mdcrd
|                           prod4.mdcrd
|
|Receptor mask:           ":1-28"
|Ligand mask:             ":29-190"
|
|Calculations performed using 50 complex frames.
|Poisson Boltzmann calculations performed using internal PBSA solver in mmpbsa_py_energy
|
|Generalized Born ESURF calculated using 'LCPO' surface areas
|
|All units are reported in kcal/mole.
```

-----  
-----  
GENERALIZED BORN:

Complex:			
Energy Component	Average	Std. Dev.	Std. Err. of Mean
-----			
VDWAALS	-1613.6541	25.0013	3.5357
EEL	-9039.4968	63.1995	8.9378
EGB	-7341.6904	58.1965	8.2302
ESURF	89.4605	0.8735	0.1235
G gas	-10653.1510	47.5106	6.7190



G solv	-7252.2298	57.4263	8.1213
TOTAL	-17905.3808	36.1286	5.1094

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-449.1800	9.3870	1.3275
EEL	4773.9608	16.9578	2.3982
EGB	-7999.2594	11.3026	1.5984
ESURF	37.1773	0.1419	0.0201
G gas	4324.7808	19.3344	2.7343
G solv	-7962.0821	11.2857	1.5960
TOTAL	-3637.3013	14.9433	2.1133

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1080.3354	12.5267	1.7715
EEL	-9352.2385	31.0148	4.3861
EGB	-3829.2975	22.5297	3.1862
ESURF	68.5386	0.3740	0.0529
G gas	-10432.5739	34.1914	4.8354
G solv	-3760.7588	22.5692	3.1918
TOTAL	-14193.3327	23.9471	3.3866

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-84.1387	11.8922	1.6818
EEL	-4461.2192	67.9124	9.6043
EGB	4486.8665	66.1993	9.3620
ESURF	-16.2554	0.6132	0.0867
DELTA G gas	-4545.3579	58.4674	8.2685
DELTA G solv	4470.6111	65.6188	9.2799
DELTA TOTAL	-74.7468	10.5830	1.4967

POISSON BOLTZMANN:

Complex:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1613.6541	25.0013	3.5357

EEL	-9039.4968	63.1995	8.9378
EPB	-7369.9665	55.3641	7.8297
ENPOLAR	65.5257	0.6917	0.0978
G gas	-10653.1510	47.5106	6.7190
G solv	-7304.4408	54.7269	7.7396
TOTAL	-17957.5918	35.4201	5.0092

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-449.1800	9.3870	1.3275
EEL	4773.9608	16.9578	2.3982
EPB	-8134.9694	11.7631	1.6636
ENPOLAR	27.5955	0.0840	0.0119
G gas	4324.7808	19.3344	2.7343
G solv	-8107.3739	11.7864	1.6669
TOTAL	-3782.5930	13.8430	1.9577

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1080.3354	12.5267	1.7715
EEL	-9352.2385	31.0148	4.3861
EPB	-3706.8826	23.2464	3.2875
ENPOLAR	50.3799	0.1848	0.0261
G gas	-10432.5739	34.1914	4.8354
G solv	-3656.5027	23.3071	3.2961
TOTAL	-14089.0765	21.7880	3.0813

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-84.1387	11.8922	1.6818
EEL	-4461.2192	67.9124	9.6043
EPB	4471.8855	70.9477	10.0335
ENPOLAR	-12.4498	0.4923	0.0696
EDISPER	0.0000	0.0000	0.0000
DELTA G gas	-4545.3579	58.4674	8.2685
DELTA G solv	4459.4357	70.4711	9.9661
DELTA TOTAL	-85.9222	15.3422	2.1697

-----  
-----

(c) Free energy output file for 1ZME.

```
| Run on Wed Jun 14 11:12:15 2017
|
| Input file:
|-----
| Input file for running PB and GB
| &general
|   endframe=50, verbose=1,
| # entropy=1,
| /
| &gb
|   igb=2, saltcon=0.100
| /
| &pb
|   istrng=0.100, inp=1, radiopt=0
| /
|-----
| MMPBSA.py Version=14.0
| Solvated complex topology file: PUT3_solvated.prmtop
| Complex topology file:      complex.prmtop
| Receptor topology file:     receptor.prmtop
| Ligand topology file:      ligand.prmtop
| Initial mdcrd(s):          density.mdcrd
|                             equil.mdcrd
|                             heat.mdcrd
|                             prod1.mdcrd
|                             prod2.mdcrd
|                             prod3.mdcrd
|                             prod4.mdcrd
|
| Receptor mask:             ":1-34"
| Ligand mask:               ":35-178"
|
| Calculations performed using 50 complex frames.
| Poisson Boltzmann calculations performed using internal PBSA solver in mmpbsa_py_energy
|
| Generalized Born ESURF calculated using 'LCPO' surface areas
|
| All units are reported in kcal/mole.
```

-----  
-----  
GENERALIZED BORN:

Complex:			
Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1578.7433	24.2172	3.4248
EEL	-6994.4162	39.5189	5.5888
EGB	-9047.4981	28.8177	4.0754
ESURF	98.5545	0.3811	0.0539
G gas	-8573.1594	39.0857	5.5276

G solv	-8948.9436	28.8017	4.0732
TOTAL	-17522.1030	36.7234	5.1935

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-551.1048	8.0432	1.1375
EEL	7225.9697	18.9110	2.6744
EGB	-10924.5689	13.3863	1.8931
ESURF	44.1960	0.1930	0.0273
G gas	6674.8649	20.7594	2.9358
G solv	-10880.3729	13.3880	1.8934
TOTAL	-4205.5080	13.3289	1.8850

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-887.7795	15.8159	2.2367
EEL	-9169.9453	31.6714	4.4790
EGB	-3235.2704	16.1739	2.2873
ESURF	75.6453	0.4331	0.0612
G gas	-10057.7248	35.4712	5.0164
G solv	-3159.6252	16.0517	2.2700
TOTAL	-13217.3500	25.6513	3.6276

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-139.8590	8.5987	1.2160
EEL	-5050.4405	25.2477	3.5706
EGB	5112.3412	27.2161	3.8489
ESURF	-21.2868	0.3032	0.0429
DELTA G gas	-5190.2995	22.6068	3.1971
DELTA G solv	5091.0544	27.0196	3.8211
DELTA TOTAL	-99.2451	12.0195	1.6998

POISSON BOLTZMANN:

Complex:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-1578.7433	24.2172	3.4248

EEL	-6994.4162	39.5189	5.5888
EPB	-9111.4754	29.1217	4.1184
ENPOLAR	74.6500	0.3228	0.0457
G gas	-8573.1594	39.0857	5.5276
G solv	-9036.8254	28.9279	4.0910
TOTAL	-17609.9848	38.6047	5.4595

Receptor:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-551.1048	8.0432	1.1375
EEL	7225.9697	18.9110	2.6744
EPB	-11085.2936	12.8364	1.8153
ENPOLAR	33.5574	0.0883	0.0125
G gas	6674.8649	20.7594	2.9358
G solv	-11051.7362	12.8536	1.8178
TOTAL	-4376.8713	13.7739	1.9479

Ligand:

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-887.7795	15.8159	2.2367
EEL	-9169.9453	31.6714	4.4790
EPB	-3151.1263	15.4184	2.1805
ENPOLAR	57.0975	0.1597	0.0226
G gas	-10057.7248	35.4712	5.0164
G solv	-3094.0288	15.4253	2.1815
TOTAL	-13151.7536	26.9650	3.8134

Differences (Complex - Receptor - Ligand):

Energy Component	Average	Std. Dev.	Std. Err. of Mean
VDWAALS	-139.8590	8.5987	1.2160
EEL	-5050.4405	25.2477	3.5706
EPB	5124.9446	28.8072	4.0740
ENPOLAR	-16.0049	0.2034	0.0288
EDISPER	0.0000	0.0000	0.0000
DELTA G gas	-5190.2995	22.6068	3.1971
DELTA G solv	5108.9396	28.6624	4.0535
DELTA TOTAL	-81.3599	14.8819	2.1046

-----  
 -----

**Appendix H:** Entropy output files calculated using *Nmode* for (a) 5EGB, (b) 1PYI, and (c) 1ZME complexes.

(a) Entropy output file for 5EGB.

```
Run on Wed Sep 28 22:54:44 2016
|
|Input file:
|-----
|Input file for running entropy calculations using NMode
|&general
|  endframe=50, keep_files=2,
|/
|&nmode
|  nmstartframe=5, nmendframe=45,
|  nminterval=5, nmode_igb=1, nmode_istrng=0.1,
|/
|
|-----
|MMPBSA.py Version=14.0
|Solvated complex topology file: 5egb_solvated.prmtop
|Complex topology file:      complex_5egb.prmtop
|Receptor topology file:    receptor_5egb.prmtop
|Ligand topology file:     ligand_5egb.prmtop
|Initial mdcrd(s):         prod1.mdcrd
|                           prod2.mdcrd
|                           prod3.mdcrd
|                           prod4.mdcrd
|
|Receptor mask:           ":1-42"
|Ligand mask:             ":43-157"
|
|Calculations performed using 50 complex frames.
|NMODE calculations performed using 9 frames.
|
|All units are reported in kcal/mole.
|All entropy results have units kcal/mol (Temperature is 298.15 K).
```

-----  
ENTROPY RESULTS (HARMONIC APPROXIMATION) CALCULATED WITH NMODE:

```
Complex:
Entropy Term      Average      Std. Dev.  Std. Err. of Mean
-----
Translational     16.7799     0.0000     0.0000
Rotational        17.3637     0.0069     0.0024
Vibrational       2287.4616   10.4491    3.6943
Total             2321.6050   10.4556    3.6966
```

```
Receptor:
Entropy Term      Average      Std. Dev.  Std. Err. of Mean
-----
```

Translational	16.1442	0.0000	0.0000
Rotational	16.6330	0.0079	0.0026
Vibrational	1032.6940	0.9374	0.3125
Total	1065.4713	0.9401	0.3134

Ligand:

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	16.1836	0.0000	0.0000
Rotational	16.6525	0.0198	0.0066
Vibrational	1339.1380	4.1335	1.3778
Total	1371.9740	4.1342	1.3781

Differences (Complex - Receptor - Ligand):

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	-15.5479	0.0000	0.0000
Rotational	-15.9218	0.0224	0.0079
Vibrational	-84.3704	11.2760	3.9867
DELTA S total=	-115.8403	11.2825	3.9890

## (b) Entropy output file for 1PYI.

```
| Run on Mon Oct 10 05:05:11 2016
|
| Input file:
|-----
| Input file for running entropy calculations using NMode
| &general
|   endframe=50, keep_files=2,
| /
| &nmode
|   nmstartframe=5, nmendframe=45,
|   nminterval=5, nmode_igb=1, nmode_istrng=0.1,
| /
|-----
| MMPBSA.py Version=14.0
| Solvated complex topology file: 1pyi_solvated.prmtop
| Complex topology file:      complex_1pyi.prmtop
| Receptor topology file:     receptor_1pyi.prmtop
| Ligand topology file:      ligand_1pyi.prmtop
| Initial mdcrd(s):          prod1.mdcrd
|                             prod2.mdcrd
|                             prod3.mdcrd
|                             prod4.mdcrd
|
| Receptor mask:             ":1-28"
| Ligand mask:               ":29-190"
```

| Calculations performed using 50 complex frames.  
| NMODE calculations performed using 9 frames.

| All units are reported in kcal/mole.  
| All entropy results have units kcal/mol (Temperature is 298.15 K).

-----  
ENTROPY RESULTS (HARMONIC APPROXIMATION) CALCULATED WITH NMODE:

Complex:  
Entropy Term            Average            Std. Dev.    Std. Err. of Mean  
-----  
Translational            16.7999            0.0000        0.0000  
Rotational                17.4651            0.0047        0.0016  
Vibrational               2463.9818           6.1703        2.0568  
Total                      2498.2469           6.1720        2.0573

Receptor:  
Entropy Term            Average            Std. Dev.    Std. Err. of Mean  
-----  
Translational            15.7799            0.0000        0.0000  
Rotational                15.7684            0.0109        0.0036  
Vibrational               681.8565            0.9926        0.3309  
Total                      713.4046            1.0006        0.3335

Ligand:  
Entropy Term            Average            Std. Dev.    Std. Err. of Mean  
-----  
Translational            16.4615            0.0000        0.0000  
Rotational                16.8775            0.0132        0.0044  
Vibrational               1818.9096           6.6770        2.2257  
Total                      1852.2484           6.6815        2.2272

Differences (Complex - Receptor - Ligand):  
Entropy Term            Average            Std. Dev.    Std. Err. of Mean  
-----  
Translational            -15.4415            0.0000        0.0000  
Rotational                -15.1808            0.0172        0.0057  
Vibrational               -36.7843            8.3988        2.7996  
  
DELTA S total=            -67.4061            8.4120        2.8040

-----  
-----  
  
(c) Entropy output file for 1ZME.

Run on Fri Jun 16 22:11:47 2017

Input file:



|Input file for running entropy calculations using NMode

|&general

| endframe=50, keep\_files=2,

|/

|&nmode

| nmstartframe=5, nmendframe=45,

| nminterval=5, nmode\_igb=1, nmode\_istrng=0.1,

|/

-----  
|MMPBSA.py Version=14.0

|Solvated complex topology file: PUT3\_solvated.prmtop

|Complex topology file: complex.prmtop

|Receptor topology file: receptor.prmtop

|Ligand topology file: ligand.prmtop

|Initial mdcrd(s): prod1.mdcrd

| prod2.mdcrd

| prod3.mdcrd

| prod4.mdcrd

|Receptor mask: ":1-34"

|Ligand mask: ":35-178"

|Calculations performed using 50 complex frames.

|NMODE calculations performed using 9 frames.

|All units are reported in kcal/mole.

|All entropy results have units kcal/mol (Temperature is 298.15 K).

-----  
ENTROPY RESULTS (HARMONIC APPROXIMATION) CALCULATED WITH NMODE:

Complex:

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	16.7900	0.0000	0.0000
Rotational	17.7585	0.0191	0.0064
Vibrational	2418.8258	3.9141	1.3047
Total	2453.3744	3.9250	1.3083

Receptor:

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	15.9552	0.0000	0.0000
Rotational	16.1807	0.0086	0.0029
Vibrational	827.9481	1.2267	0.4089
Total	860.0842	1.2316	0.4105

Ligand:

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	16.3502	0.0000	0.0000
Rotational	17.4263	0.0312	0.0104
Vibrational	1642.9734	5.8529	1.9510
Total	1676.7499	5.8529	1.9510

Differences (Complex - Receptor - Ligand):

Entropy Term	Average	Std. Dev.	Std. Err. of Mean
Translational	-15.5154	0.0000	0.0000
Rotational	-15.8485	0.0378	0.0126
Vibrational	-52.0957	9.0660	3.0220
DELTA S total=	-83.4597	9.0717	3.0239

-----  
-----

## REFERENCES

- (1) Ganguly, A.; Rajdev, P.; Williams, S. M.; Chatterji, D. Nonspecific Interaction between DNA and Protein Allows for Cooperativity: A Case Study with Mycobacterium DNA Binding Protein. *J. Phys. Chem. B* **2012**, *116* (1), 621–632.
- (2) Clark, P. *Molecular Biology*, 2nd ed.; Elsevier: London, 2013.
- (3) Harrison, S., A structural taxonomy of DNA-binding domains. *Nature*. **1991**, *353*, 715-716.
- (4) Bujn icki, J., Prediction of Protein Structures, Functions, and Interactions; John Wiley and Sons: United Kindom, 2009.
- (5) Green, A.; Parker, M.; Conte, D.; Sarkar, B. Zinc Finger Proteins: A Bridge between Transition Metals and Gene Regulation. *J. Trace Elem. Exp. Med.* **1998**, *11* (2–3), 103–118.
- (6) Fu, F.; Sander, J. D.; Maeder, M.; Thibodeau-Beganny, S.; Joung, J. K.; Dobbs, D.; Miller, L.; Voytas, D. F. Zinc Finger Database (ZiFDB): A Repository for Information on C2H2 Zinc Fingers and Engineered Zinc-Finger Arrays. *Nucleic Acids Res.* **2009**, *37* (SUPPL. 1), 279–283.
- (7) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc Finger Proteins : New Insights into Structural and Functional Diversity. *J. Struct. Biol.* **2001**, *11*, 39–46.
- (8) Iuchi,Sh.; Kludell, N., *Zinc Finger Proteins: From Atomic Contact to Cellular Function*; Plenum:USA, 2005.
- (9) Malone, R., *Bioinorganic Chemistry*, 2<sup>nd</sup> ed.; John Wiley and Sons: New Jersey, 2007.
- (10) Jacobs, G. H. Determination of the Base Recognition Positions of Zinc Fingers from Sequence Analysis. *EMBO J.* **1992**, *11* (12), 4507–4517.
- (11) Pellegrino, G. R.; Berg, J. M. Identification and Characterization Of “zinc-Finger” domains by the Polymerase Chain Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (2), 671–675.
- (12) Pabo, C.; Peisach, E.; Grant, R. Design and Selection of Novel Cyc2His2 Zinc Finger Proteins. *Annu Rev Biochem* **2001**, *70* (November), 313–340.
- (13) Krishna, S. S.; Majumdar, I.; Grishin, N. V. Structural Classification of Zinc

- Fingers. *Nucleic Acids Res.* **2003**, *31* (2), 532–550.
- (14) Bouhouche, N.; Syvanen, M.; Kado, C. I. The Origin of Prokaryotic C<sub>2</sub>H<sub>2</sub> Zinc Finger Regulators. *Trends Microbiol.* **2000**, *8* (2), 77–81.
  - (15) Searles, M. a; Lu, D.; Klug, A. The Role of the Central Zinc Fingers of Transcription Factor IIIA in Binding to 5 S RNA. *J. Mol. Biol.* **2000**, *301* (1), 47–60.
  - (16) Wolfe, S.; Nekludova, L.; Pabo, C., Dna Recognition By Cys<sub>2</sub>His<sub>2</sub> Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, 183–212.
  - (17) Brayer, K. J.; Segal, D. J. Keep Your Fingers off My DNA: Protein-Protein Interactions Mediated by C<sub>2</sub>H<sub>2</sub> Zinc Finger Domains. *Cell Biochem. Biophys.* **2008**, *50* (3), 111–131.
  - (18) Lachenmann, M. J.; Ladbury, J. E.; Phillips, N. B.; Narayana, N.; Qian, X.; Weiss, M. a. The Hidden Thermodynamics of a Zinc Finger. *J. Mol. Biol.* **2002**, *316*, 969–989.
  - (19) Sarisky, C. A.; Mayo, S. L. The B $\beta$  $\alpha$  Fold: Explorations in Sequence Space. *J. Mol. Biol.* **2001**, *307* (5), 1411–1418.
  - (20) Matthews, J. M.; Sunde, M. Zinc Fingers — Folds for Many Occasions. *IUMB Life* **2002**, *54*, 351–355.
  - (21) Wuttke, D. S.; Foster, M. P.; Case, D. a; Gottesfeld, J. M.; Wright, P. E. Solution Structure of the First Three Zinc Fingers of TFIIIA Bound to the Cognate DNA Sequence: Determinants of Affinity and Sequence Specificity. *J. Mol. Biol.* **1997**, *273* (1), 183–206.
  - (22) Pavletich, N. P.; Pabo, C. Zinc Structure of a Recognition : Complex Zif268-DNA. *Adv. Sci.* **1991**, *252* (5007), 809–817.
  - (23) Miller, J.; McLachlan, A. D.; Klug, A. Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from *Xenopus* Oocytes. *J. Trace Elem. Exp. Med.* **2001**, *14* (2), 157–169.
  - (24) Baker, C. L.; Petkova, P.; Walker, M.; Flachs, P.; Mihola, O.; Trachtulec, Z.; Petkov, P. M.; Paigen, K. Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. *PLoS Genet.* **2015**, *11* (9), 1–24.
  - (25) Berg, I. L.; Neumann, R.; Lam, K.-W. G.; Sarbajna, S.; Odenthal-Hesse, L.; May, C. a; Jeffreys, A. J. PRDM9 Variation Strongly Influences Recombination Hot-Spot Activity and Meiotic Instability in Humans. *Nat. Genet.* **2010**, *42* (10), 859–863.

- (26) Patel, A.; Horton, J. R.; Wilson, G. G.; Zhang, X.; Cheng, X. Structural Basis for Human PRDM9 Actions at Recombination Hotspots SM. *Genes Dev.* **2016**, *30* (3), 257–265.
- (27) Striedner, Y.; Schwarz, T.; Welte, T.; Futschik, A.; Rant, U.; Tiemann-Boege, I. The Long Zinc Finger Domain of PRDM9 Forms a Highly Stable and Long-Lived Complex with Its DNA Recognition Sequence. *Chromosom. Res.* **2017**.
- (28) Berg, I. L.; Neumann, R.; Lam, K.-W. G.; Sarbajna, S.; Odenthal-Hesse, L.; May, C. a; Jeffreys, A. J. PRDM9 Variation Strongly Influences Recombination Hot-Spot Activity and Meiotic Instability in Humans. *Nat. Genet.* **2010**, *42* (10), 859–863.
- (29) MacPherson, S.; Larochele, M.; Turcotte, B. A Fungal Family of Transcriptional Regulators: The Zinc Cluster Proteins. *Microbiol. Mol. Biol. Rev.* **2006**, *70* (3), 583–604.
- (30) Schjerling, P.; Holmberg, S. Comparative Amino Acid Sequence Analysis of the C6 Zinc Cluster Family of Transcriptional Regulators. *Nucleic Acids Res.* **1996**, *24* (23), 4599–4607.
- (31) Correll, C.; Rice, P., *Protein-Nucleic Acid Interactions: Structural Biology*; RSC: Cambridge, 2008.
- (32) Campbell, R. N.; Leverentz, M. K.; Ryan, L. a; Reece, R. J. Metabolic Control of Transcription: Paradigms and Lessons from *Saccharomyces Cerevisiae*. *Biochem. J.* **2008**, *414* (2), 177–187.
- (33) Turcotte, B.; Liang, X. B.; Robert, F.; Soontornngun, N. Transcriptional Regulation of Nonfermentable Carbon Utilization in Budding Yeast. *FEMS Yeast Res.* **2010**, *10* (1), 2–13.
- (34) Ljungdahl, P. O.; Daignan-Fornier, B. Regulation of Amino Acid, Nucleotide, and Phosphate Metabolism in *Saccharomyces Cerevisiae*. *Genetics* **2012**, *190* (3), 885–929.
- (35) Maret, W.; Larsen, K. S.; Vallee, B. L. Coordination Dynamics of Biological Zinc “clusters” in Metallothioneins and in the DNA-Binding Domain of the Transcription Factor Gal4. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (6), 2233–2237.
- (36) Zhang, L.; Guarente, L. The C6 Zinc Cluster Dictates Asymmetric Binding by HAP1. *EMBO J.* **1996**, *15* (17), 4676–4681.

- (37) Ball, L. J.; Diakun, G. P.; Gadhavi, P. L.; Young, N. A.; Armstrong, E. M.; David Garner, C.; Laue, E. D. Zinc Co-Ordination in the DNA-Binding Domain of the Yeast Transcriptional Activator PPR1. *FEBS Lett.* **1995**, 358 (3), 278–282.
- (38) Marmorstein, R.; Harrison, S., Crystal Structure of a PPR1-DNA Complex : DNA Recognition by Proteins Containing a Zn<sub>2</sub>Cys<sub>6</sub> binuclear Cluster. *Genes & Development*, **1994**, 2504–2512.
- (39) Flynn, P. J.; Reece, R. J. Activation of Transcription by Metabolic Intermediates of the Pyrimidine Biosynthetic Pathway. *Mol. Cell. Biol.* **1999**, 19 (1), 882–888.
- (40) Liang, S. D.; Marmorstein, R.; Harrison, S. C.; Ptashne, M. DNA Sequence Preferences of GAL4 and PPR1: How a Subset of Zn<sub>2</sub> Cys<sub>6</sub> Binuclear Cluster Proteins Recognizes DNA. *Mol. Cell. Biol.* **1996**, 16 (7), 3773–3780.
- (41) Xu, S.; Falvey, D. a; Brandriss, M. C. Roles of URE2 and GLN3 in the Proline Utilization Pathway in *Saccharomyces Cerevisiae*. *Mol. Cell. Biol.* **1995**, 15 (4), 2321–2330.
- (42) Spitzner, A.; Perzlmaier, A. F.; Geillinger, K. E.; Reihl, P.; Stolz, J. The Proline-Dependent Transcription Factor Put3 Regulates the Expression of the Riboflavin Transporter MCH5 in *Saccharomyces Cerevisiae*. *Genetics* **2008**, 180 (4), 2007–2017.
- (43) Axelrod, J. D.; Majors, J.; Brandriss, M. C. Proline-Independent Binding of PUT3 Transcriptional Activator Protein Detected by Footprinting in Vivo. *Mol. Cell. Biol.* **1991**, 11 (1), 564–567.
- (44) Swaminathan, K.; Flynn, P.; Reece, R.; Marmorstein, R., Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn<sub>2</sub>Cys<sub>6</sub> binuclear cluster. *Nature Struct. Bio.*, **1997**, 4 (9), 751- 759.
- (45) Jen, J.; Wang, Y.-C. Zinc Finger Proteins in Cancer Progression. *J. Biomed. Sci.* **2016**, 23 (1), 53.
- (46) Jamieson, A. C.; Miller, J. C.; Pabo, C. O. Drug Discovery with Engineered Zinc-Finger Proteins. *Nat. Rev. Drug Discov.* **2003**, 2 (5), 361–368.
- (47) Durai, S.; Mani, M.; Kandavelou, K.; Wu, J. Zinc Finger Nucleases as Gene Therapy Agents. *Nucleic acids* **2005**, 15 (22), 1463–1468.
- (48) Miller, J., An improved zinc-finger nuclease architecture for highly specific

genome editing. *Nature Biotechnol*, **2007**, 25(7): 778–785

- (49) Xiao, Y.; Xiang, T.; Luo, X.; Li, C.; Li, Q.; Peng, W.; Li, L.; Li, S.; Wang, Z.; Tang, L.; et al. Zinc-Finger Protein 545 Inhibits Cell Proliferation as a Tumor Suppressor through Inducing Apoptosis and Is Disrupted by Promoter Methylation in Breast Cancer. *PLoS One* **2014**, 9 (10), 1–8.
- (50) Swanson, J. M. J.; Henchman, R. H.; Mccammon, J. A. Revisiting Free Energy Calculations : A Theoretical Connection to MM / PBSA and Direct Calculation of the Association Free Energy. *Biophysical journal*, **2004**, 86, 67–74.
- (51) Hayes, J.; Archonites, G., Molecular Dynamics: Studies of Synthetic and Biological Macromolecules. InTech: London, 2011.
- (52) Beierlein, F. R.; Kneale, G. G.; Clark, T. Predicting the Effects of Basepair Mutations in DNA-Protein Complexes by Thermodynamic Integration. *Biophysj* **2011**, 101 (5), 1130–1138.
- (53) Yunta, M. J. . How to Calculate Binding Constants for Drug Discovery Studies. *Am. J. Model. Optim.* **2013**, 1 (3), 61–70.
- (54) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson À Boltzmann Surface Area Method. *MOL. Inf.*, **2012**, 114–122.
- (55) Yunta, M. J. . How to Calculate Binding Constants for Drug Discovery Studies. *Am. J. Model. Optim.* **2013**, 1 (3), 61–70.
- (56) Amber, C. Amber 2016 Reference Manual. **2016**.
- (57) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Donini, O.; Cieplak, P.; et al. Calculating Structures and Free Energies of Complex Molecules : Combining Molecular Mechanics and Continuum Models. **2000**, 33 (12), 889–897.
- (58) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, 2 (1), 128–139.
- (59) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, 23 (14), 1297–1304.
- (60) Hayes, J. M.; Archontis, G. MM-GB ( PB ) SA Calculations of Protein-

Ligand Binding Free Energies. **2011**.

- (61) Honig, B.; Nicholls, A.; Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. **2007**, 268 (5214), 1144–1149.
- (62) Allen, M., Introduction to Molecular Dynamics Simulation. *NIC: New york*, **2004**, 23, 1-28.
- (63) Nurisso, A.; Daina, A.; Walker, R. C. *Homology Modeling*; 2012; Vol. 857.
- (64) Maruani, J., Molecules in Physics, Chemistry, and Biology: Electronic Structure and Chemical Reactivity. *Kluwer Academic:London*, 2012.
- (65) Lee, J.; Kim, J.-S.; Seok, C. Cooperativity and Specificity of Cys 2 His 2 Zinc Finger Protein–DNA Interactions: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2010**, 114 (22), 7662–7671.
- (66) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins. *Structure* **2000**, 8 (10), 1015–1023.
- (67) Privalov, P. L.; Dragan, A. I.; Crane-Robinson, C.; Breslauer, K. J.; Remeta, D. P.; Minetti, C. A. S. A. What Drives Proteins into the Major or Minor Grooves of DNA? *J. Mol. Biol.* **2007**, 365 (1), 1–9.
- (68) Hamed, M. Y.; Arya, G. Zinc Finger Protein Binding to DNA: An Energy Perspective Using Molecular Dynamics Simulation and Free Energy Calculations on Mutants of Both Zinc Finger Domains and Their Specific DNA Bases. *J. Biomol. Struct. Dyn.* **2016**, 34 (5), 919–934.
- (69) Sousa, F.; Cruz, C.; Queiroz, J. A. Amino Acids-Nucleotides Biomolecular Recognition: From Biological Occurrence to Affinity Chromatography. *J. Mol. Recognit.* **2010**, 23 (6), 505–518.
- (70) Hoffman, M. M.; Khrapov, M. A.; Cox, J. C.; Yao, J.; Tong, L.; Ellington, A. D. AANT: The Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res.* **2004**, 32 (DATABASE ISS.), 174–181.
- (71) Hou, T.; Wang, J.; Li, Y.; Wang, W.; Houa, T.; Wangb, J.; Lia, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods: I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Comput. Sci.* **2011**, 51 (1), 69–82.
- (72) Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols



- Using PDBbind Data Set. *Phys. Chem. Chem. Phys.* **2014**, *16* (31), 16719.
- (73) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM / PBSA and MM / GBSA Methods . I . The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model* **2011**, *51*, 69–82.
- (74) Rezácová, P.; Borek, D.; Moy, S. F.; Joachimiak, A.; Otwinowski, Z. Crystal Structure and Putative Function of Small Toprim Domain-Containing Protein from *Bacillus Stearotherophilus*. *Proteins* **2008**, *70* (2), 311–319.
- (75) Wang, J.; Wang, W.; Kollman, P. a; Case, D. a. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *J. Am. Chem. Soc* **2001**, *222* (2), U403.
- (76) Rother, K. Introduction to PyMOL. *Methods Mol. Biol.* , **2005**, *635* (8), 0–32.
- (77) (<http://www.rbvi.ucsf.edu/chimera/1.8/docs/UsersGuide/index.html> [6/4/13 1:04:51 PM]).